THE TRENDS IN EFFICIENCY OF LITHUANIAN DAIRY FARMS: A SEMIPARAMETRIC APPROACH

Tomas Baležentis¹, Tianxiang Li², Alvydas Baležentis³

¹ Dr. Lithuanian Institute of Agrarian Economics. Address: Kudirkos Str. 18-2, 03105 Vilnius, Lithuania. Tel. +370 52 621078. E-mail tomas@laei.lt
² PhD candidate. Nanjing Agricultural University. E-mail garylee0428@gmail.com
³ Prof. Dr. Mykolas Romeris University. E-mail a.balezentis@gmail.com

Received 13 04 2015; accepted 13 05 2015

This study aims at analysing the trends in efficiency of Lithuanian dairy farms and thus identifying the prospective development paths. The semiparametric approach based on nonparametric regression and Stochastic Frontier Analysis is applied for the analysis. The research relies on Farm Accountancy Data Network and covers family farms. The period of 2004–2011 is considered. In order to identify the underlying trends in dairy farming, we focus on such features as technical efficiency, partial elasticities, and elasticity of scale. The semiparametric approach yielded rather high efficiencies. Specifically, the average technical efficiency of 89% was observed. A decline in technical efficiency during 2004–2011 is present for both point estimates and associated bounds of the confidence interval. Analysis of the elasticity of scale implies that most of the farms could still increase their scale of operation. The obtained results were confirmed by a parametric random coefficients model.

Keywords: dairy farms, efficiency, semiparametric analysis, production frontier, stochastic frontier analysis.

JEL codes: C14, D24, Q12.

1. Introduction

Efficiency analysis is facilitated by the means of frontier techniques. These can be grouped into parametric and nonparametric ones. The key difference between these two broad categories is that the former ones require imposition of a specific functional form upon a representation of the underlying technology (e.g., production function, cost function, profit function), whereas the latter ones do not require such-like assumptions. The analysis proceeds by considering the gaps among observations (production plans) and the production frontier estimated by either of the frontier methods. Given different frontier techniques might render different representations of the technology, different patterns of efficiency scores might be revealed. Therefore, it is important to apply different techniques to analyse the dynamics in efficiency and productivity. For a wider discussion in regards to frontier methodology, please refer to, for instance, L. R. Murillo-Zamorano (2004) and R. Färe et al. (2013).
Data Envelopment Analysis (DEA), as defined by A. Charnes et al. (1978), is the main nonparametric method. In spite of appealing features of empirical frontier satisfying conditions of monotonicity, convexity, and minimal extrapolation (Afriat, 1972) pertinent to DEA, the latter technique is a deterministic one. This implies that the whole gap in productivity is explained by inefficiency. Stochastic Frontier Analysis (SFA), introduced independently by D. Aigner et al. (1977) and W. Meeusen and J. van den Broeck (1977), can be considered as the main parametric technique. Being a stochastic method, SFA allows for isolation of the statistical noise within the composite error term thus allowing for random deviations from a representation of the technology. The trade-off involved is an a priori assumption on the functional form of a representation of the technology. The estimated frontier might not satisfy such desirable properties as monotonicity and convexity. See J. S. Liu et al. (2011) and C. F. Parmeter and S. C. Kumbhakar (2014) for surveys on extensions of DEA and SFA, respectively.

There have been various attempts to introduce stochasticity in nonparametric framework by considering chance-constrained programming (Land, 1993) or fuzzy set theory (Hatami-Marbini, 2011), among other approaches. On the other hand, A. Henningsen and C. H. Henning (2009) offered a framework to impose restrictions of monotonicity on the stochastic frontiers. Yet another remedy for the aforementioned shortcomings is the framework introduced by Y. Fan et al. (1996), unifying certain benefits of nonparametric and parametric techniques. Specifically, the proposed approach relies upon nonparametric regression alongside SFA. Therefore, the expected output value (in case of production function) is estimated without any specific assumptions regarding the functional form the production frontier, yet the resulting error terms is decomposed into random error and inefficiency term in the spirit of SFA.

Agricultural sector requires analysis of its efficiency due to the fact that competition is not fully enforced there due to public support and non-farm activities. Under these circumstances, suchlike analysis might reveal possible ways for improvement in performance and effectiveness of public support schemes. Lithuanian dairying sector features rather intensive transformations fuelled by changes in input and output prices. Indeed, the share of milk – the main output of dairy farms – in the national gross agricultural output dropped from 21.8% in 2009 down to 16.5% in 2013. The literature regarding efficiency of dairy farms in Lithuania is rather limited. For instance, A. Jedik et al. (2014) analysed the efficiency of Lithuanian dairy farms by means of the deterministic parametric frontier. Specifically, the translog production frontier was assumed. As the European Union has entered the programming period of 2014–2020, there is a need to streamline strategic decisions regarding farm specialisation, farm structure, and public support through the Common Agricultural Policy. Accordingly, it is important to deliver scientific evidence regarding the prospective developments in the aforementioned areas.

This study therefore aims at analysing the trends in efficiency of Lithuanian dairy farms. The semiparametric approach of Y. Fan et al. (1996) is applied for the analysis. Specifically, the research relies on Farm Accountancy Data Network (FADN) and covers family farms. The period of 2004–2011 is considered. In order to
identify the underlying trends in dairy farming, we focus on such features as technical efficiency, partial elasticities, and elasticity of scale.

The paper proceeds as follows: Section 2 presents the semiparametric framework for efficiency analysis. Section 3 describes data used for the analysis. Section 4 brings the results.

2. Preliminaries for a semiparametric analysis of efficiency

The semiparametric setting of Y. Fan et al. (1996) relies on nonparametric regression and SFA. In the first stage, nonparametric regression is employed to define a production frontier. This stage is devoid of any assumptions on the functional form of the production frontier. In the second stage, SFA is employed to decompose the resulting error terms. For this, certain assumptions are imposed over the distributions of the terms of the composite error term. This section heavily relies on O. Badunenko et al. (2012).

Assume that each farm (indexed by \( k = 1, 2, \ldots, K \)) consumes the vector of inputs \( x_k \in \mathbb{R}^p \) to produce output quantity \( y_k \in \mathbb{R}_+ \). The output is then related to the vector of inputs as follows:

\[
y_k = g(x_k) + v_k - u_k, \quad k = 1, 2, \ldots, K, \tag{1}
\]

where \( g(\cdot) \) is an unknown smooth function to be estimated nonparametrically, \( v_k \sim N(0, \sigma^2_v) \) is a random error term, \( u_k \sim N(0, \sigma^2_u) \) is an inefficiency term.

The unknown production function, \( g(\cdot) \), can be estimated via nonparametric regression. Specifically, such estimators as local-constant least squares or local-linear least squares can be employed (cf. Henderson, 2015) to obtain estimates of interest. However, one cannot observe \( g(\cdot) \) directly, as the nonparametric regression yields an estimate of the following expectation:

\[
\hat{E}(y_k | x_k) = g(x_k) + E(u_k | x_k).
\]

In case of local-linear least squares regression, the following estimator is applied to obtain \( \hat{E}(y_k | x_k) \):

\[
\min_{\alpha, \beta} \sum_{k=1}^{K} (y_k - \alpha - (x_k - x) \beta)^2 K_h(x, x_k), \tag{2}
\]

where \( \hat{E}(y_k | x_k) = \alpha \), \( \alpha \) is a scalar, \( x \) and \( \beta \) are vectors of appropriate dimensions, and \( K_h(\cdot, \cdot) \) is the kernel function with bandwidth \( h \). Package \( np \) (Hayfield, 2008) is employed to implement the nonparametric regression.

Given both random error and inefficiency term are identically and independently distributed, the results of the first stage can be passed to SFA for further decomposition. Specifically, we assume normal distribution of the random error along with half-normal distribution of the inefficiency term. Following D. Aigner et al. (1997), the following likelihood function is maximized over \( \lambda \):
\[ I(\lambda) = -K \ln \hat{\sigma} + \sum_{k=1}^{K} \ln \left( 1 - \Phi \left( \frac{\lambda \hat{\varepsilon}_k}{\hat{\sigma}} \right) \right) - \frac{1}{2\hat{\sigma}^2} \sum_{k=1}^{K} \hat{\varepsilon}_k^2, \quad (3) \]

where \( \hat{\varepsilon}_k = y_k - \hat{E}(y_k | x_k) - \mu(\hat{\sigma}, \lambda) \) with

\[ \mu(\hat{\sigma}, \lambda) = \left( \sqrt{2\lambda \hat{\sigma}} / \left( \pi(1 + \lambda^2) \right) \right)^{1/2}, \quad \text{and} \quad (4) \]

\[ \hat{\sigma}(\lambda) = \left[ \frac{1}{K} \sum_{k=1}^{K} (y_k - \hat{E}(y_k | x_k))^2 / \left( 1 - \frac{2\lambda^2}{\pi(1 + \lambda^2)} \right) \right]^{1/2}. \quad (5) \]

Given \( \lambda = \sigma_u / \sigma_v \) and \( \sigma^2 = \sigma_u^2 + \sigma_v^2 \), one can recover variances of the error term and inefficiency. These enable to estimate inefficiency term as (Jondrow, 1982):

\[ E(u_k | \varepsilon_k) = \mu_u + \sigma_u \left[ \frac{\phi(-\mu_u / \sigma_u)}{1 - \Phi(-\mu_u / \sigma_u)} \right], \quad (6) \]

with \( \mu_u = \varepsilon_u \sigma_u^2 / (\sigma_u^2 + \sigma_v^2) \) and \( \sigma^2 = \sigma_u^2 \sigma_v^2 / (\sigma_u^2 + \sigma_v^2) \). The technical efficiency is then given as \( \exp(-u_k) | \varepsilon_k \).

The confidence intervals for the point estimator given by Eq. 6 can be estimated in lines with W. C. Horrace and P. Schmidt (1996). Specifically, the lower and upper boundaries of \( \exp(-u_k) | \varepsilon_k \) for confidence level of \((1 - \alpha)100\%\) are defined as:

\[ L_k = \exp(-\mu_u - z_{1\alpha} \sigma_u), \quad \text{and} \]

\[ U_k = \exp(-\mu_u - z_{1\alpha} \sigma_u), \quad (7) \]

where

\[ z_{1\alpha} = \Phi^{-1} \left( 1 - \frac{\alpha}{2} \left( 1 - \Phi \left( \frac{-\mu_u}{\sigma_u} \right) \right) \right), \quad \text{and} \quad (9) \]

\[ z_{1\alpha} = \Phi^{-1} \left( 1 - \frac{\alpha}{2} \left( 1 - \Phi \left( \frac{-\mu_u}{\sigma_u} \right) \right) \right). \quad (10) \]

The presented approach allows for a flexible production function, yet it does not guarantee such properties as monotonicity. We use the code provided by O. Badunenko et al. (2012) to implement the approach of Fan et al. (1996). In order to check the robustness of the results, we employ the parametric Cobb-Douglas model. Specifically, random coefficients panel model is applied for the comparison.
3. Data used

The FADN data cover the period of 2004–2011. Note that the farm classification has changed in the meantime. Therefore, for dairy farms we look at farming type 41 under regulation of 2003 for the period of 2004–2009 and farming type 45 under regulation of 2008 for 2010 and onwards.

The four inputs and one output were considered to model the technology. Output is total agricultural output. Labour input is measured in annual work units (AWU). Utilised agricultural area in ha is treated as land input. Intermediate consumption includes specific costs and overheads. Finally, asset value less the value of land is used as a capital input. Table 1 below summarises the input-output variables and their development throughout the time. The total number of observations is 1832.

It is evident that dairy farms have expanded in Lithuania (the data presented here, however, are not weighted). The highest rate of growth, viz. 136%, was observed for asset input. This has been induced by public support and successive investments into modern equipment. Intermediate consumption has increased by 57%. The lowest rates of growth have been observed for labour and land inputs (12–15%). As the total output has increased by 42%, which is lower than the growth rates associated with some of the inputs, one can expect to observe an increasing inefficiency in Lithuanian dairying farms.

Table 1. Mean values of inputs and outputs, 2004–2011

<table>
<thead>
<tr>
<th>Year</th>
<th>Total output (thousand LTL)</th>
<th>Labour input (AWU)</th>
<th>Land area (ha)</th>
<th>Intermediate consumption (thousand LTL)</th>
<th>Assets (thousand LTL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>183</td>
<td>2.4</td>
<td>81.9</td>
<td>95</td>
<td>331</td>
</tr>
<tr>
<td>2005</td>
<td>219</td>
<td>2.5</td>
<td>80.8</td>
<td>108</td>
<td>448</td>
</tr>
<tr>
<td>2006</td>
<td>218</td>
<td>2.6</td>
<td>78.5</td>
<td>113</td>
<td>552</td>
</tr>
<tr>
<td>2007</td>
<td>217</td>
<td>2.6</td>
<td>72.2</td>
<td>107</td>
<td>560</td>
</tr>
<tr>
<td>2008</td>
<td>256</td>
<td>2.9</td>
<td>86.8</td>
<td>139</td>
<td>740</td>
</tr>
<tr>
<td>2009</td>
<td>194</td>
<td>2.7</td>
<td>81.5</td>
<td>119</td>
<td>727</td>
</tr>
<tr>
<td>2010</td>
<td>240</td>
<td>2.8</td>
<td>92.3</td>
<td>133</td>
<td>763</td>
</tr>
<tr>
<td>2011</td>
<td>261</td>
<td>2.8</td>
<td>91.8</td>
<td>149</td>
<td>783</td>
</tr>
<tr>
<td>Sample average</td>
<td>227</td>
<td>2.7</td>
<td>84.7</td>
<td>123</td>
<td>634</td>
</tr>
</tbody>
</table>

Monetary variables were deflated by respective real price indices with base year 2005. To avoid negative values of the gradients, a minimum of 1.01 AWU was assumed for the labour input. Note that 1 EUR equals 3.4528 LTL. Inputs and outputs enter the model in a logged form thus enabling to estimate partial elasticities associated with each of inputs as well as elasticity of scale.
4. Results

The approach of Y. Fan et al. (1996) yielded rather high efficiencies, for non-parametric regression renders an extremely flexible production frontier. Specifically, the average technical efficiency (TE) of 89% was observed. This implies that output could have been increased by some 11% on average. As regards the trends in efficiency, a decline in TE is present for both point estimates and associated bounds of the confidence interval. As one can note, the upper bound of TE scores is rather high and is probably related to well-fitted production frontier.

As input-output variables entered the model in their log form, gradient of the nonparametric frontier is interpreted as a vector of partial elasticities. The descriptive statistics for these estimates are given in Table 3.

Note that negative values are observed for some instances as the production frontier does not satisfy conditions of monotonicity globally. The highest value of partial elasticity was observed for intermediate consumption (0.85). For asset input, the corresponding average value was much lower (0.11).

<table>
<thead>
<tr>
<th>Year</th>
<th>Lower bound</th>
<th>Point estimate</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>0.76</td>
<td>0.90</td>
<td>0.995</td>
</tr>
<tr>
<td>2005</td>
<td>0.76</td>
<td>0.90</td>
<td>0.995</td>
</tr>
<tr>
<td>2006</td>
<td>0.74</td>
<td>0.89</td>
<td>0.994</td>
</tr>
<tr>
<td>2007</td>
<td>0.75</td>
<td>0.90</td>
<td>0.995</td>
</tr>
<tr>
<td>2008</td>
<td>0.73</td>
<td>0.88</td>
<td>0.993</td>
</tr>
<tr>
<td>2009</td>
<td>0.70</td>
<td>0.86</td>
<td>0.989</td>
</tr>
<tr>
<td>2010</td>
<td>0.73</td>
<td>0.88</td>
<td>0.992</td>
</tr>
<tr>
<td>2011</td>
<td>0.73</td>
<td>0.88</td>
<td>0.993</td>
</tr>
<tr>
<td>Sample average</td>
<td>0.74</td>
<td>0.89</td>
<td>0.993</td>
</tr>
</tbody>
</table>

Labour and land input featured extremely low partial elasticities. These findings imply that intermediate consumption plays the most important role in the production process. Therefore, improvements in fodder etc. are still important in Lithuanian dairy farms.

Table 3. Descriptive statistics for partial elasticities, 2004–2011

<table>
<thead>
<tr>
<th>Minimum</th>
<th>Labour input</th>
<th>Land area</th>
<th>Intermediate consumption</th>
<th>Assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06</td>
<td>0.06</td>
<td>0.02</td>
<td>0.85</td>
<td>0.11</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.85</td>
<td>0.66</td>
<td>1.14</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Partial elasticities are summed up to estimate elasticity of scale. If elasticity of scale is greater (lower) than unity indicates, then increasing (decreasing) returns to
scale prevails. Kernel density estimate for elasticity of scale is depicted in Fig. 1. The average value of scale elasticity was 1.05, yet Fig. 1 suggests that most of the observations were concentrated in the region of increasing RTS. This implies that most of the farms could still increase their scale of operation.

Indeed, smaller farms face lower milk selling prices. Even though we use price indices for deflation of agricultural output, farm-specific differences in selling prices remain unaccounted for. Therefore, increase in scale of the operation might not only induce improvements in input utilisation in technological sense, but also improve marketing efficiency. As regards milk selling prices, agricultural cooperatives should take measures to ensure higher prices for their members more actively.

Fig. 1. Kernel density for scale elasticity, 2004–2011

The estimates of scale elasticity are related to total output and land input to obtain some insights into the optimal dairy farm size. Fig. 2 exhibits the relationship between total output and scale elasticity. The trend implies that the most productive scale size is maintained at the output level of some 500 thousand LTL, which is twice larger than the average value reported in Table 1. Therefore, it is important to increase both yield and price of production through animal recording and other relevant practices besides expansion of the productive activities.

As Fig. 3 suggests, the optimal farm should employ some 234 ha for agricultural production. Indeed, the latter figure is also much higher than sample average. Finally, Fig. 4 implies that the optimal farm size is achieved when assets amount to some 2.2 million LTL. Note, however, that these estimates are based on inefficient observations and thus could be reduced by some 11% on average if full technical efficiency were achieved. It should also be stressed that increase in farm size is related to additional adjustment costs. Therefore, farm expansion might lead into increasing efficiency in the long run.

Comparison of Figs. 2–4 suggests that the sample farms are specific with a rather compact distribution of total output values and land input. However, one can observe that there is a substantial share of dairy farms with vast amounts of assets. The latter finding
stresses again the impact of excessive investments and adjustment costs upon dairy farm performance in Lithuania.

Fig. 2. Relationship between total output and scale elasticity

\[ y = 1.6086x^{-0.036} \]
\[ R^2 = 0.2453 \]

Fig. 3. Relationship between utilised agricultural area and scale elasticity

\[ y = 1.1973x^{-0.033} \]
\[ R^2 = 0.1543 \]

Fig. 4. Relationship between asset input and scale elasticity

\[ y = 1.5051x^{-0.028} \]
\[ R^2 = 0.1608 \]
Even though this study supposes that farm expansion is preferable, some caveats should be considered. Indeed, whenever farm growth is fuelled by public support, large farms are those which benefit the most. Therefore, support measures should aim at supporting small- and mid-size farms at first.

In order to check the robustness of results, a mixed linear model is estimated assuming Cobb-Douglas production function. Input quantities are treated as fixed effects, whereas farm and time dummies are included as random effects. The estimates of the coefficients of the production function are presented in Table 4.

Table 4. Estimates of the Cobb-Douglas model with random effects

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.9763</td>
<td>0.1315</td>
<td>7.4</td>
</tr>
<tr>
<td>log(labour)</td>
<td>0.0662</td>
<td>0.0158</td>
<td>4.2</td>
</tr>
<tr>
<td>log(land)</td>
<td>0.0883</td>
<td>0.0154</td>
<td>5.7</td>
</tr>
<tr>
<td>log(intermediate)</td>
<td>0.7245</td>
<td>0.0176</td>
<td>41.1</td>
</tr>
<tr>
<td>log(assets)</td>
<td>0.1793</td>
<td>0.0121</td>
<td>14.8</td>
</tr>
</tbody>
</table>

If compared to results based on nonparametric regression, the partial elasticity associated with intermediate consumption became somewhat lower, whereas those associated with land and asset inputs increased. The sum of partial elasticities is 1.06 indicating increasing returns to scale.

Fig. 5. Kernel density for technical efficiency based on Cobb-Douglas production function

The estimates of the random effects are normalised as \( \exp(e_k - \max_k e_k) \) for each \( k \) (similarly, the same procedure is applied to time effects) to obtain farm-specific and time-specific efficiencies. Fig. 5 presents a stochastic kernel for the resulting efficiencies. In this case, we analyse farm-specific TE score, which are constant throughout the time.
The mean farm-specific TE was 0.54, which is much smaller figure if opposed to that based on the approach of Y. Fan et al. (1996). The latter difference might be induced by a more restrictive functional form of the Cobb-Douglas frontier. Furthermore, TE scores depicted in Fig. 5 neglect temporal developments of efficiency. The latter part of efficiency is presented in Fig. 6.

![Fig. 6. The trend in technical efficiency based on random effects associated with time periods, 2004–2011](image)

As Fig. 6 suggests, the TE followed a downward trend in general during 2004–2011. Years 2006 and 2009 mark especially steep decreases in efficiency due to unfavourable climatic conditions. The general trend might be related to various factors. First, the growth in input prices might not be followed by substantial growth in output quantity and prices. Second, a vibrant accumulation of assets (cf. Table 1) is related to adjustment costs.

Comparison of the approach of Y. Fan et al. (1996) and Cobb-Douglas production frontier shows that both efficiency trends and partial elasticities (as well as elasticity of scale) are similar across the two settings. However, the differences in mean levels of efficiency are evident possibly due to different functional forms of the production frontiers. Anyway, the results can be considered as being robust in qualitative sense under both of the approaches.

5. Conclusions

1. The sample data show that dairy farms have expanded in Lithuania. The highest rate of growth, viz. 136%, was observed for asset input. This has been induced by public support and successive investments into modern equipment. Intermediate consumption has increased by 57%. The lowest rates of growth have been observed for labour and land inputs (12–15%). As the total output has increased by 42%, which is lower than the growth rates associated with some of the inputs, one can expect to observe an increasing inefficiency in Lithuanian dairying farms.
2. The semiparametric approach yielded rather high efficiencies. Specifically, the average technical efficiency of 89% was observed. This implies that output could have been increased by some 11% on average. A decline in technical efficiency during 2004–2011 is present for both point estimates and associated bounds of the confidence interval.

3. The average value of scale elasticity was 1.05, and kernel density suggested that most of the observations were concentrated in the region of increasing returns to scale. This implies that most of the farms could still increase their scale of operation.

4. The agricultural policy in Lithuania should aim at increasing efficiency of dairy farms via modernisation, which could also benefit in terms of farm expansion. Given the relatively low livestock grazing intensity in Lithuania, suchlike developments seem to be feasible. Furthermore, efficiency analysis should be conducted for homogeneous groups of farms of different size in order to ascertain whether especial measures are needed for farms of particular size.

5. The obtained results were confirmed by a parametric random coefficients model. Further research should focus on a constrained estimation of the semiparametric production frontier (Martins-Filho, Yao, 2015) in order to ensure regularity conditions. In addition, panel settings can be exploited to account for farm heterogeneity. Contextual variables might be included in the analysis to identify the main factors of inefficiency.

Acknowledgements

Tianxiang Li would like to thank for the support of China Scholarship Council and China Institute for Rural Studies, Tsinghua University.

References


LIETUVOS PIENININKYSTĖS ŪKIŲ EFEKTYVUMO TENDENCIJOS: SEMIPARAMETRINIS POŽIŪRIS

Tomas Baležentis¹, Tianxiang Li², Alvydas Baležentis³

¹ Lietuvos agrarinės ekonomikos institutas; ² Nanjing žemės ūkio universitetas; ³ Mykolo Romerio universitetas

Įteikta 2015 04 13; priimta 2015 05 13

Santrauka


Raktiniai žodžiai: pienininkystės ūkiai, efektyvumas, semiparametrinė analizė, gamybos riba, stochastinė ribų analizė.

JEL koda: C14, D24, Q12.