Proceedings of the 2015 International Conference "ECONOMIC SCIENCE FOR RURAL DEVELOPMENT" No37
Jelgava, LLU ESAF, 23-24 April 2015, pp. 35-44

THE IMPACT OF TIME SERIES EXPANSION IN NON-PARAMETRIC ANALYSES OF EFFICIENCY EFFECTS

Tomas Balezentis[+], Dr
Lithuanian Institute of Agrarian Economics

Abstract. Efficiency analyses often involve panel data which enable to reveal certain longitudinal patterns. However, it is often impossible to maintain the same sample structure when expanding time series. This paper, thus, attempts to test whether there are obvious differences in the trends of efficiency across two different data sets. The research focuses on Lithuanian family farm performance. The data are taken form Farm Accountancy Data Network with different samples covering the periods of 2004-2009 and 2004-2011. The juxtaposition of two-stage efficiency analyses based on different data sets enabled to identify the impact of changes in both the structure of the sample and time series expansion upon the efficiency factors and effects. The analysis suggests that Lithuanian family farms featured generally the same patterns of efficiency during 2004-2011 if opposed to 2004-2009.

Key words: data envelopment analysis, efficiency, family farms, Lithuania.

JEL code: C14, C34, C44, D24, Q12.

Introduction

Efficiency and productivity gains are the most important factors behind a sustainable economic growth. Indeed, gains in efficiency enable to allocate the resources in a more rational way across the economy. Therefore, it is important to develop reasonable frameworks for analysis of efficiency and productivity. As regards the agricultural sector, it is due to Gorton and Davidova (2004) and Balezentis (2014b) that the frontier methods are the most widely applied ones for the efficiency analyses. Efficiency analysis requires estimation of the productive technology, which can be described in terms of the production function, distance function, cost function, or revenue function, among others. The two wide groups of the frontier methods can be delineated as regards specification of the representation of the productive technology, viz., parametric methods and non-parametric methods (Coelli et al., 2005). The parametric methods require a priori specification of the functional form of the representation, whereas the non-parametric methods do not require suchlike assumptions. These

[+] Corresponding author. Tel.: + 370 5 262 10 78; fax: + 370 5 261 45 24
E-mail address: tomas@laei.lt

circumstances make the non-parametric methods quite appealing ones if opposed to the parametric ones. Yet, the non-parametric methods do not allow for analysis of efficiency effects. On the contrary, parametric methods, like Stochastic Frontier Analysis (SFA), proposed by Meeusen and van den Broeck (1977) and Aigner et al. (1977), can accommodate the efficiency effects (Kumbhakar et al., 1991; Battese, Coelli, 1995). Data Envelopment Analysis (DEA), introduced by Charnes et al. (1978) and Banker et al. (1984), does not allow for inclusion of the efficiency effects into the model. Therefore, two-stage analysis is conducted to estimate the impact of certain factors upon the DEA-based efficiency scores. Initially, such techniques as the Ordinary Least Squares and Tobit model were used for suchlike analyses. It is due to Simar and Wilson (2007) that the aforementioned setting might render misleading results due to the underlying serial correlation among the observations. Accordingly, the double bootstrap methodology has been proposed to overcome these issues. Indeed, the double bootstrap methodology can be considered as semi-parametric one, as the second stage analysis relies on the truncated regression model. Daraio and Simar (2005, 2007a, 2007b) introduced the conditional measures of efficiency. The latter framework allows for a fully non-parametric analysis of the efficiency factors. The proposed model assumes no separability among environmental variables and the shape of the underlying technology set. Fousekis et al. (2014) employed the said framework in the agricultural context. Balezentis et al. (2014) suggested a fully non-parametric framework for analysis of the efficiency effects assuming separability. In the latter case, no conditional measures are involved.

Analyses of the agricultural efficiency do often rely on the data from Farm Accountancy Data Network (FADN) or other statistical data bases. Accordingly, the data are not easily and readily available for the research. On the contrary, multiple aspects of the efficiency should be analysed by employing the same datasets in order to avoid additional sampling bias. Therefore, one often needs to check whether one data set features the same underlying trends as opposed to those associated with another data set.

The aim of the research is to propose a procedure for testing whether there are obvious differences in the trends of efficiency across two different data sets. The following tasks are set: 1) to present the preliminaries of frontier-based efficiency analysis; 2) to present the data available for analysis of Lithuanian family farm performance; 3) to analyse the impact of time series expansion on the results of efficiency analysis. This paper employs the approach proposed by Balezentis et al. (2014) to analyse the differences in efficiency effects due to expansion of the time span of the analysis. Particularly, the paper focuses on the two datasets, each covering the periods of 2004-2009 and 2004-2011. These data are taken form Farm Accountancy Data Network (Lithuanian Institute of Agrarian Economics, 2012). In addition, the results are compared with those obtained via the double bootstrap methodology.

Methodological approach

The fully non-parametric framework proposed by Balezentis et al. (2014) relies on bootstrapped DEA and non-parametric regression. The bootstrapped DEA is described by Simar and Wilson (1998) and implemented by Wilson (2008). The obtained efficiency scores are then used as dependent variables in the non-parametric regression (Racine, Li, 2004; Li, Racine, 2007) as it is implemented by Hayfield and Racine (2008). In its essence, the non-parametric regression weights the residuals at each observation with respect to its distances to other observations as defined by the kernel function when optimising the sum of squared residuals. The proposed framework uses partial regression plots to depict the relationships among the efficiency factors and efficiency scores.

As regards the double bootstrap methodology, the present research follows Algorithm #2 in Simar and Wilson (2007). The latter technique comprises the two parts. First, the bootstrap technical efficiency scores are estimated in order to approximate the true underlying technology with respect to the environmental variables. Second, the bootstrap coefficients of the truncated regression are estimated to establish the corresponding confidence intervals.

In order to maintain comparability with the previous results, the input-oriented model was employed for methodology proposed by Balezentis et al. (2014), whereas the output-oriented model was applied for the double bootstrap (Simar, Wilson, 2007).

Research results and discussion

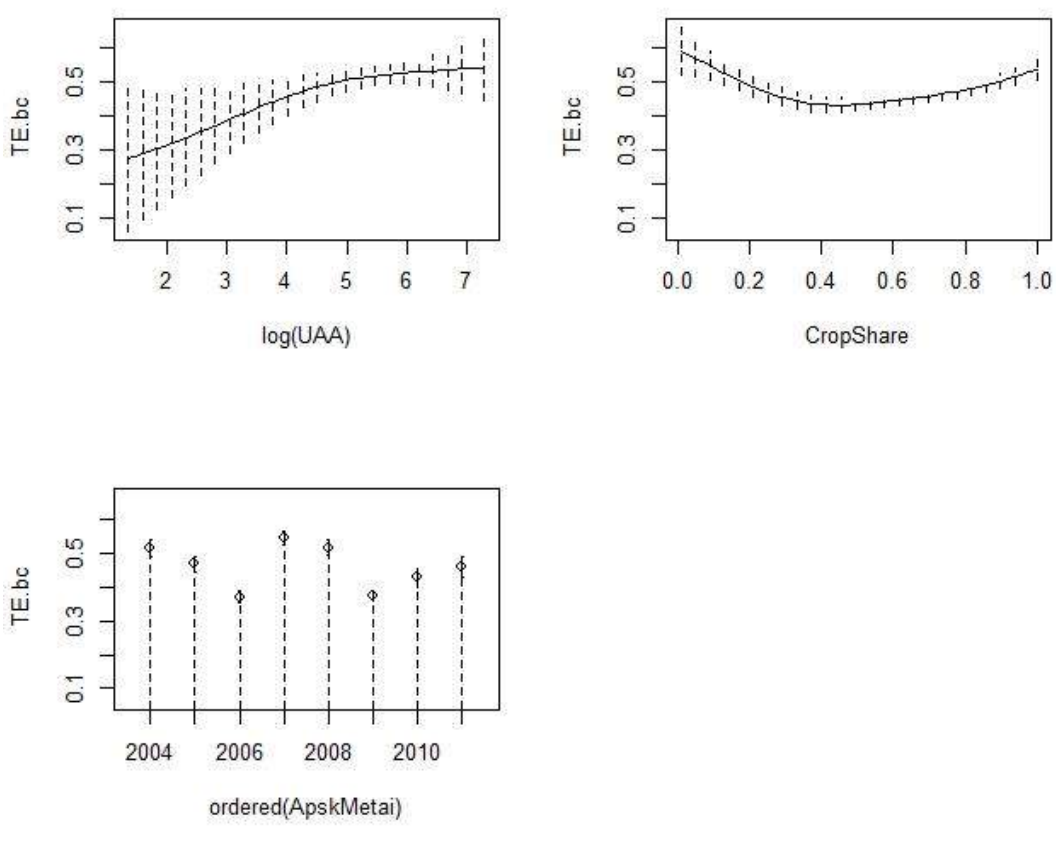
Balezentis et al. (2014) and Balezentis (2014a) employed the FADN data set for years 2004-2009. The aforementioned dataset is a balanced panel comprising 200 observations (viz., family farms) per year. Indeed, the end of the said period coincides with economic turmoil. This paper, thus, attempts to check the consistency of the obtained results by fitting models used in the aforementioned research to the extended data set.

In order to proceed with the comparison, the additional dataset was extracted from the FADN. The new dataset covers the period of 2004-2011. Given the FADN farm sample changes year to year, the expansion of time span renders a decrease in the number of farms observed for each year to 163. The extended dataset, thus, is a balanced panel comprising 1304 observations in total. The third dataset has also been established for the same 163 farms, yet covering the initial time span.

To sum up, the following three samples were considered: 1) years 2004-2009, 200 farms; 2) years 2004-2011, 163 farms; 3) years 2004-2009, 163 farms. Samples 1 and 3 share the same time period albeit they cover different farms, hence, they can be employed to test for differences in efficiency effects due to changes in the sample structure. Samples 2 and 3 share the same sample structure yet cover different time spans, thus, they can be used to test for differences in efficiency effects arising due to temporal developments (possibly, those occurring in the whole family farming sector).

The technical efficiency was estimated by considering four inputs, namely utilised agricultural land area in hectares, labour input in annual work units, intermediate consumption in Litas, and assets in Litas, and one output, viz., total agricultural output in Litas. The monetary variables were deflated by respective real price indices. Various explanatory variables were chosen for the analysis, yet only significant ones were finally kept in lines of the backward procedure.

Results of the non-parametric regression are presented in Figure 1. As one can note, results based on the dataset for 2004-2011 and 163 farms virtually re-iterate those obtained by considering the original data set for 2004-2009 and 200 farms (Balezentis et al., 2014). Specifically, technical efficiency appeared to increase with increasing farm size, yet a kink in the partial regression plot was observed at the (logged) value of 400 ha. The time trend showed decreases in technical efficiency during 2006 and 2009. The only evident difference between the original and extended datasets is an increase in crop farm efficiency. Anyway, livestock farming appear to be the most efficient on average. These differences might have been caused by both changes in relative performance and sample structure. Therefore, the third sample was further considered.



Source: author's calculations based on Lithuanian Institute of Agrarian Economics (2012).

Fig. 1. **Partial regression plots (2004-2011, 163 farms, the dependent variable is Farrell input efficiency)**

The following Table 1 presents the bandwidths and p-values. Obviously, all the variables were statistically significant at 1%.

Table 1

Results of non-parametric regression analysis (2004-2011, 163 farms)

	log(*UAA*)	*CropShare*	ordered(*Year*)
Bandwidth	1.447169	0.1780017	0.04444178
P Value	<.000 ***	<.000 ***	<.000 ***

*Significance codes: *** – 0.001, ** – 0.01, and * – 0.05.*

Source: author's calculations based on Lithuanian Institute of Agrarian Economics (2012).

The double bootstrap analysis was also re-iterated with the extended data set. The resulting truncated regression's coefficients are given in Table 2. The new results can be compared against those in Balezentis (2014). Note that more variables entered the model under the backward procedure. These include time trend (*Trend*) to check if there is an underlying general trend describing the change in efficiency, utilised agricultural area (*UAA*) to account for farm size, the ratio of assets to the labour input (*Assets/AWU*) to measure the level of capital accumulation and modernisation (to a certain extent), the ratio of subsidies over the total output (*Subsidies*) to measure the impact of production subsidies on the farm performance. Note that the confidence intervals were defined for the three levels of confidence by the means of bias-corrected accelerated and percentiles methods.

Table 2

Double bootstrap estimates for determinants of the farming inefficiency (2004-2011, 163 farms)

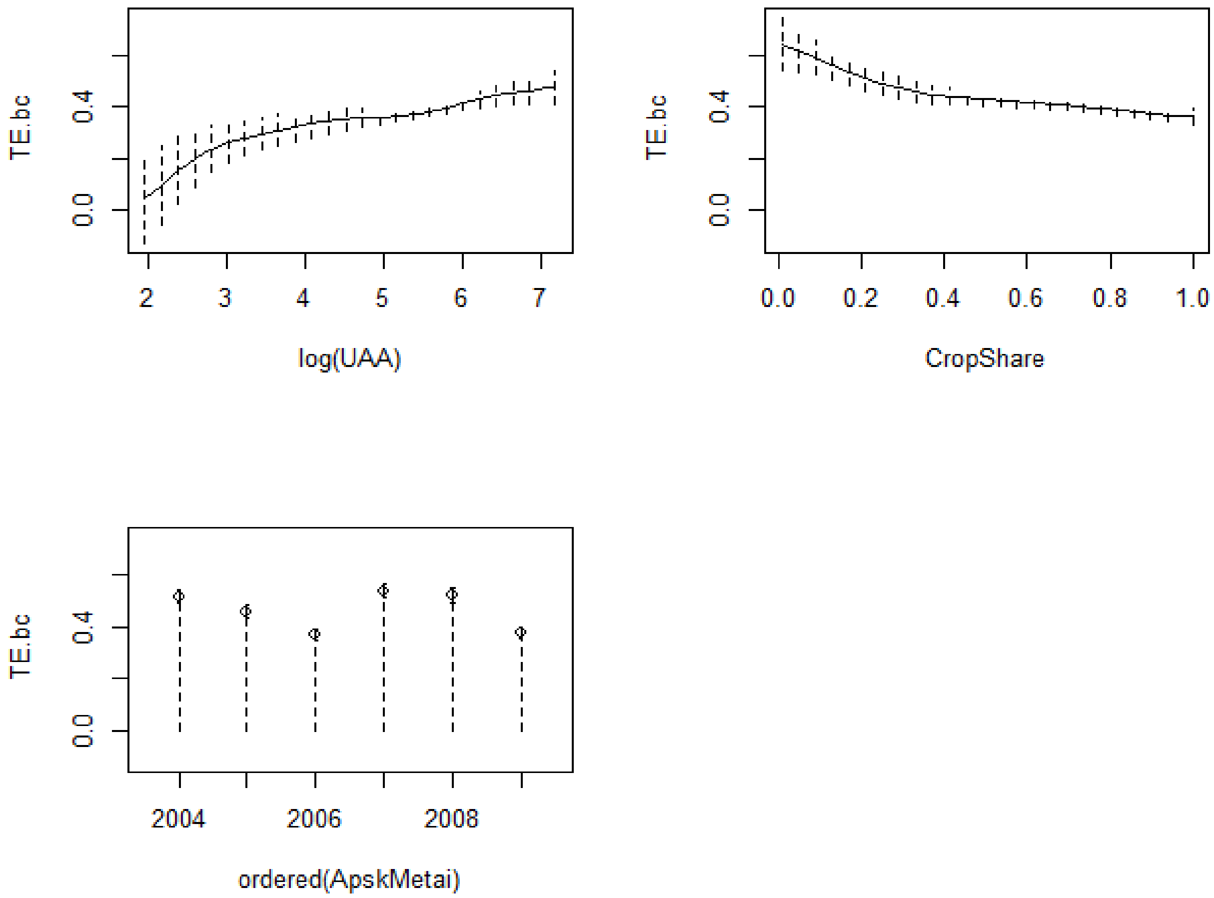
Variables	$\hat{\hat{\beta}}$	**Sig.**	**Confidence intervals**					
			$\alpha = .1$		$\alpha = .05$		$\alpha = .01$	
			BC_a **method**					
Time	0.175	***	0.114	0.240	0.103	0.256	0.082	0.280
UAA	-0.073		-0.240	0.095	-0.278	0.122	-0.331	0.177
Assets/AWU	-0.167	*	-0.361	-0.007	-0.397	0.024	-0.506	0.078
Crop	0.714	***	0.321	1.051	0.249	1.113	0.068	1.225
Subsidies	1.520	***	1.394	1.649	1.372	1.678	1.331	1.730
			Percentiles method					
Time	0.175	***	0.111	0.237	0.100	0.251	0.076	0.276
UAA	-0.073		-0.242	0.094	-0.278	0.119	-0.331	0.176
Assets/AWU	-0.167		-0.347	0.006	-0.381	0.032	-0.483	0.091
Crop	0.714	***	0.339	1.066	0.274	1.134	0.086	1.236
Subsidies	1.520	***	1.392	1.646	1.370	1.676	1.322	1.721

*Significance codes: '***' - 0.01, '**' - 0.05, '*' - 0.1; the dependent variable is Farrell output efficiency*

Source: author's calculations based on Lithuanian Institute of Agrarian Economics (2012).

Results obtained for the extended data set do not contradict to those based on the original data set presented by Balezentis (2014). The only significant difference is a change in the direction of the time trend: the extended data set suggests a negative time trend. However, this can be a direct outcome of expansion of timespan. Farm size features a positive effect upon efficiency, yet the associated coefficient is no longer significant. The asset-labour ratio is also specific with the same direction of the relationship. However, it is insignificant according to the percentiles method. The remaining two variables, viz. crop share in the total output and production subsidy intensity, featured the same kind of relationships with the efficiency. These findings are also supported by results of the fully non-parametric framework (cf. Figure 1).

As it was already said, certain differences emerged between results presented by Balezentis (2014) and Balezentis et al. (2014) might be due to changes in the structure of the sample under analysis. Therefore, the new sample is reduced by considering only the period of 2004-2009 and keeping the same 163 farms in the sample. The resulting estimates of the coefficients of the non-parametric regression are depicted in Figure 2.



Source: author's calculations based on Lithuanian Institute of Agrarian Economics (2012).

Fig. 2. **Partial regression plots (2004-2009, 163 farms, the dependent variable is Farrell input efficiency)**

Table 3 presents the p-values associated with the regressors. Evidently, all the three variables were significant.

Table 3

Results of non-parametric regression analysis (2004-2011, 163 farms)

	log(*UAA*)	*CropShare*	ordered(*Year*)
Bandwidth	0.6636057	0.1894315	0.05482321
P Value	<.000 ***	<.000 ***	<.000 ***

*Significance codes: *** – 0.001, ** – 0.01, and * – 0.05.*

Source: author's calculations based on Lithuanian Institute of Agrarian Economics (2012).

The results of double bootstrap model for the 163 farms during 2004-2009 are presented in Table 4. If compared to results for the same period based on the larger data set for the 200 farms, it is evident that the change in the sample structure played an important role. Therefore, the impact of farm size (UAA) and assets per AWU became insignificant. Crop output's share in the total output had a negative effect upon efficiency, yet the latter factor remained insignificant at the level of significance of 10% under one of the methods for estimation of the confidence intervals (anyway, the lower bound of the confidence interval suggests that p-value might not be far from 0.1).

Table 4

Double bootstrap estimates for determinants of the farming inefficiency (2004-2009, 163 farms)

Variables	$\hat{\hat{\beta}}$	**Sig.**	**Confidence intervals**					
			$\alpha = .1$		$\alpha = .05$		$\alpha = .01$	
			BC_a **method**					
Time	0.345	***	0.250	0.452	0.226	0.470	0.189	0.512
UAA	-0.020		-0.219	0.179	-0.253	0.215	-0.328	0.283
Assets/AWU	-0.135		-0.353	0.042	-0.404	0.066	-0.476	0.123
Crop	0.493		-0.026	0.871	-0.110	0.953	-0.324	1.101
Subsidies	1.347	***	1.215	1.481	1.195	1.505	1.148	1.556
			Percentiles method					
Time	0.345	***	0.245	0.451	0.224	0.468	0.184	0.509
UAA	-0.020		-0.232	0.169	-0.270	0.207	-0.339	0.279
Assets/AWU	-0.135		-0.341	0.051	-0.391	0.075	-0.472	0.138
Crop	0.493	*	0.074	0.932	-0.050	0.988	-0.231	1.161
Subsidies	1.347	***	1.210	1.478	1.192	1.502	1.145	1.555

*Significance codes: '***' - 0.01, '**' - 0.05, '*' - 0.1; the dependent variable is Farrell output efficiency*

Source: author's calculations based on Lithuanian Institute of Agrarian Economics (2012).

Results based on the sample of 163 farms and the time period of 2004-2009 did indicate that the change in structure of the sample (while keeping the time period fixed) induced null effects of farm size in terms of UAA and the ratio of assets to labour input. However, the latter variable became significant at the confidence level of 10% after time series expansion. Obviously, time series expansion resulted in decrease in the coefficient associated with the time trend, thus, implying a positive development of family farms' efficiency during 2010-2011. These results are also confirmed by a fully non-parametric framework (cf. Figures 1 and 2). Even though the farm size variable (*UAA*) remained insignificant after time series expansion, the coefficient as well as its upper confidence band decreased, thus, implying that larger farms gained more productivity if compared to smaller ones. Nevertheless, the coefficient remained insignificant at the confidence level of 10%. The negative impact of production subsidies persisted and even increased during the period of 2010-2011.

In order to obtain more precise estimates of the p-values in the fully non-parametric framework, Daraio and Simar (2014) introduced an improved bootstrapping methodology. In addition, the test proposed by Daraio et al. (2010) might be utilised to check whether the condition of separability holds for the analysed data. Finally, Badin et al. (2012) argued that the conditional framework enables to estimate "pure" and managerial efficiency.

Conclusions

1. The juxtaposition of two-stage efficiency analyses based on different data sets enabled to identify the impact of changes in both the structure of the sample and time series expansion upon the efficiency factors and effects. The analysis suggests that Lithuanian family farms featured generally the same patterns of efficiency during 2004-2011 if opposed to 2004-2009.
2. Crop farms became more efficient during 2010-2011 as opposed to 2004-2009. The carried out analysis also revealed that a part of the latter increase was due to changes in the structure of the analysed sample. Indeed, this finding is in line with the undergoing changes in Lithuanian family farms, where more and more farmers opt for crop farming. Anyway, it was livestock farms that maintained the highest efficiency.
3. The negative impact of production subsidies persisted and even increased during the period of 2010-2011. Farm size effect became insignificant in the double bootstrap setting due to changes in the sample structure, yet certain changes were observed after time series expansion.
4. For sake of comparison, the present research utilised models with exactly the same variables as it was the case with the original data set. Therefore, further analyses should attempt to apply the methodologies proposed in this paper with the extended data set in order to reveal a possible impact of inclusion of additional variables into analysis.
5. As regards the non-parametric analysis of the efficiency effects, further researches should attempt to test the condition of separability, analyse managerial efficiency in the conditional settings, and perform the improved tests of significance of the efficiency factors.

Bibliography

1. Aigner, D., Lovell, C. A. A., Schmidt, P. (1977). Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics*, Volume 6, Issue 1, pp. 21-37.
2. Badin, L., Daraio, C., Simar, L. (2012). How to Measure the Impact of Environmental Factors in a Nonparametric Production Model. *European Journal of Operational Research*, Volume 223, Issue 3, pp. 818-833.
3. Balezentis, T. (2014a). Analysing the Determinants of Lithuanian Family Farm Performance: A Double Bootstrap Inference. *Economic Science for Rural Development*, Volume 34, pp. 66–74.
4. Balezentis, T. (2014b). On Measures of the Agricultural Efficiency – A Review. *Transformations in Business & Economics*, Volume 13, Issue 3, pp. 110-131.
5. Balezentis, T., Krisciukaitiene, I., Balezentis, A. (2014). A Nonparametric Analysis of the Determinants of Family Farm Efficiency Dynamics in Lithuania. *Agricultural Economics*, Volume 45, Issue 5, pp. 489-499.
6. Banker, R. D., Charnes, A., Cooper, W. W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, Volume 30, Issue 9, pp. 1078-1092.
7. Battese, G. E., Coelli, T. J. (1995). A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics*, Volume 20, pp. 325-332.
8. Charnes, A., Cooper, W. W., Rhodes, E. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*, Volume 2, Issue 6, pp. 429-444.
9. Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. Springer. p. 349.
10. Daraio, C., Simar, L. (2005). Introducing Environmental Variables in Nonparametric Frontier Models: A Probabilistic Approach. *Journal of Productivity Analysis*, Volume 24, Issue 1, pp. 93-121.
11. Daraio, C., Simar, L. (2007a). Conditional Nonparametric Frontier Models for Convex and Nonconvex Technologies: A Unifying Approach. *Journal of Productivity Analysis*, Volume 28, Issue 1-2, pp. 13-32.
12. Daraio, C., Simar, L. (2007b). *Advanced Robust and Nonparametric Methods in Efficiency Analysis*. Springer. p. 248.
13. Daraio, C., Simar, L. (2014). Directional Distances and Their Robust Versions: Computational and Testing Issues. *European Journal of Operational Research*, Volume 237, Issue 1, pp. 358-369.
14. Daraio, C., Simar, L., Wilson, P. W. (2010). Testing Whether Two-Stage Estimation is Meaningful in Non-Parametric Models of Production (Vol. 1030). Discussion Paper. Retrieved: http://sites.uclouvain.be/IAP-Stat-Phase-V-VI/ISBApub/dp2010/DP1031.pdf. Access: 15.12.2014
15. Fousekis, P., Kourtesi, S., Polymeros, A. (2014). Assessing Managerial Efficiency on Olive Farms in Greece. *Outlook on Agriculture*, Volume 43, Issue 2, pp. 123-129.
16. Gorton, M., Davidova, S. (2004). Farm Productivity and Efficiency in the CEE Applicant Countries: A Synthesis of Results. *Agricultural Economics*, Volume 30, pp. 1-16.
17. Hayfield, T., Racine, J. S. (2008). Nonparametric Econometrics: The np Package. *Journal of Statistical Software*, Volume 27, Issue 5, pp. 1-32.

18. Kumbhakar, S.C., Ghosh, S., McGuckin, J. T. (1991). A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms. *Journal of Business and Economic Statistics*. Volume 9, pp. 279-286.

19. Li, Q., Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton: Princeton University Press. p. 768.

20. Lithuanian Institute of Agrarian Economics. (2012). *Ukiu veiklos rezultatai (UADT tyrimo duomenys) 2011* [FADN Survey Results 2011]. Vilnius: Lietuvos agrarines ekonomikos institutas. p. 108.

21. Meeusen, W., van den Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review*, Volume 18, Issue 2, pp. 435-444.

22. Racine, J. S., Li, Q. (2004) Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data. *Journal of Econometrics*, Volume 119, pp. 99-130.

23. Simar, L., Wilson, P. W. (1998). Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. *Management Science*, Volume 44, Issue 1, pp. 49-61.

24. Simar, L., Wilson, P. W. (2007). Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes. *Journal of Econometrics*, Volume 136, Issue 1, pp. 31-64.

25. Wilson, P. (2008). FEAR: A Software Package for Frontier Efficiency Analysis with R. *Socio-Economic Planning Sciences*, Volume 42, Issue 4, pp. 247-254.