Primary submission: 18.03.2018 | Final acceptance: 13.07.2018

Financial Performance Measurement of Hungarian Retail Food Companies

Veronika Fenyves[1], Zoltán Bács[1], Laura Karnai[2], Adrián Nagy[2], Tibor Tarnóczi[1]

ABSTRACT

The comparison of company performances, i.e., benchmarking, is becoming more and more critical. Presently, companies mostly use traditional financial ratios to evaluate their financial performance. We also use financial ratios to measure and compare company performances, from which we create complex efficiency coefficients using Data Envelopment Analysis. Using Data Envelopment Analysis, we analyzed the efficiency of retail food companies in Hungary's Northern Great Plain region from 2009 to 2014 using their financial reports. To improve the result of the performance measurement, we used the bootstrap method, the Hamiltonian Monte Carlo simulation, and Bayesian statistics. We transformed the primarily deterministic DEA method into a stochastic DEA model. The primary target of this extension is to enhance statistical inference in DEA and to integrate it with a stochastic mechanism of Bayesian techniques. To develop the stochastic DEA model, we use Stan Stochastic Modelling Language within the framework of the R Statistics. Analyzing the results, we can state that the DEA method can be used for analyzing efficiency, and the additions shown can make the evaluation much more accurate. We can conclude that the best results were produced by the combined method, during a simultaneous application of the input orientation.

KEY WORDS: performance measurement, principal components analysis, data envelopment analysis, bayesian statistics, stan stochastic programming language

JEL Classification: M21, L25, L81, G32, C1

[1] University of Debrecen, Faculty of Economics and Business, Institute of Accounting and Finance, Hungary
[2] University of Debrecen, Faculty of Economics and Business, Institute of Business Administration, Hungary

1. Introduction

Company leaders are under enormous pressure to increase their company's performance. Performance assessment or benchmarking is a widely used method for corporate development and profitability enhancement. Benchmarking is becoming an increasingly important activity for companies in today's globalized world. Business enterprises need to perform their operations efficiently to survive the market competition and achieve the expected level of profitability. Efficiency is a significant indicator in the companies' activities analysis, and it is one of the fundamental and most frequently used performance indicator. To measure, monitor and improve efficiency is a vital task for enterprises in the 21st century (Andrejic, Bojovic, & Kilibarda, 2013).

It is essential for every company to be able to place itself in the corporate rankings of its industry. Determining the ranking of a company within industry

Correspondence concerning this article should be addressed to: **Veronika Fenyves**, University of Debrecen, Faculty of Economics and Business, Institute of Accounting and Finance, Egyetem tér 1., Debrecen 4032. E-mail: fenyves.veronika@econ.unideb.hu

requires to measure corporate performance using a composite indicator. Company experts attempt to measure company performance using scores assigned to corporate financial indicators in general. However, to transform financial ratios into a composite indicator is quite complicated. One of the ways to create corporate industry rankings can be to use Data Envelopment Analysis (DEA), which creates a company ranking using multiple performance indicators. The DEA method allows using various input and output variables to determine a performance indicator (efficiency coefficient). However, it is essential to bear in mind that accurate performance measurement is possible only if the conditions for calculating the DEA are met (Sarkar, 2017).

The primary goal of this study is to use the DEA method to rank and compare the retail food companies of the Northern Great Plain region while defining a complex performance indicator that allows an accurate assessment of the investigated companies. The objective of the research was to achieve a correct assessment by extending the DEA method using the principal component analysis (PCA) and the Stan probabilistic programming language. Principal component analysis can help us to reduce the dimensions of input and output variables. PCA is a data reduction tool that removes redundant information, shows hidden features and essential relationships existing between observations (Putri, Chetchotsak, & Jani, 2017). The classical DEA model cannot be applied in our case because it does not have sufficient discriminatory power.

We chose the food retail sector as the subject of analysis because it is an important sector within the Hungarian economy and no similar study has been made regarding the Northern Great Plan region in the past.

The rest of this paper is organized as follows:

- The section on 'Methodology' presents the database investigated and the methods used: data envelopment analysis, principal component analysis and Stan stochastical programming language based on Bayesian statistics.
- The next section 'Results', presents the results and discussion of the analysis.
  - *Checking the number of variables and the outliers.*
  - *Determining the efficiency coefficients using PCA-DEA.*
  - *Determining the interval of DEA efficiency coefficients using Stan stochastic modeling tool.*
- Conclusions are given in the last section.

2. Methodology

2.1. Investigation database

The enterprises that were included in the study database are situated in the Northern Great Plain region of Hungary and are selected if their main activity was indicated as a "Food retail mixed store" which was established before January 1, 2009, and which have six financial years closed with an annual report (2009-2014). The enterprises were selected from the OPTEN company database, and their annual reports were downloaded from the Hungarian Electronic Reporting Portal (e-beszamolo). There were 887 companies fitted the conditions determined above in the region, 563 of which are part of the database analyzed. During the period examined, 86 enterprises were liquidated, and 238 enterprises did not produce annual reports in several years of the period examined, or the items needed to the analysis of the annual report contained zero values. The number of enterprises involved in the investigation in the region, were, by county:

- Szabolcs-Szatmár-Bereg County 131
- Hajdú-Bihar County 250
- Jász-Nagykun-Szolnok County 182
- Total Northern Great Plain Region 563

The great majority of enterprises in the database prepared a simplified annual report, and so the analysis was performed upon the data found in those simplified reports. For the analysis, we used the accounting items shown in Table 1 as input and output variables.

2.2. Data envelopment analysis

In this study, Data Envelopment Analysis was used to measure corporate performance, which was developed from mathematical programming used in operations research. DEA is, in essence, a linear programming model which allows us to rank the enterprises (decision-making units – DMU) involved in the examination by a joint utilization of different input and output characteristics. Cook and Zhu (2005) comment in their study that DEA is a data-oriented

Table 1. Accounting parameters used for performance measurement

Input variables		Output variables	
Number	**Name of variables**	**Number**	**Name of variables**
1	Fixed assets	1	Revenue
2	Inventories	2	Operating profit
3	Receivables	3	Profit before tax
4	Cash	4	Profit after tax
5	Equity		
6	Reserves		
7	Non-current liabilities		
8	Current liabilities		
9	Material costs		
10	Employee costs		
11	Depreciation		
12	Other costs		
13	Financial expenses		
14	Corporate taxes		

approach to performance evaluation and improvement and they then present the various authors' research on DEA. In several professional contexts, DEA models are also called nonparametric deterministic models, where the word 'deterministic' can be applied only to the basic model and its extensions because it is also possible to create stochastic DEA models (Huang & Li, 2011; L. Li & M. Li, 2014). This study does not deal in detail with the DEA method because many authors had dealt with it during the past 60 years what was described by Farell initially (1957). However, the real "birth" of the method is calculated from 1978 when Charnes and his co-authors published their DEA article (Charnes, Cooper, & Rhodes, 1978). Thousands of publications have been published on DEA since 1978 (Tavares 2002; Emrouznejad, Parker, & Tavares, 2008).

DEA is a useful tool for evaluating the relative performance efficiency of companies (Putri et al., 2017). The advantage of the DEA method is that it can take into account more input and output variables at the same time, using a relative efficiency measurement coefficient. Efficiency is defined as the proportion of outputs to inputs. The DEA method focuses on limit values instead of central tendencies. Researchers from various research fields quickly realized that DEA is an excellent method to model operational processes in any segment of the economy (Cooper, Seiford, & Tone, 2007).

When using DEA, the following considerations should be made:

- DEA is sensitive to low numbers of investigated units;
- DEA is sensitive to outliers;
- DEA is sensitive to a large number of input and output variables;
- DEA is sensitive to the degree of statistical noise (measurement errors).

2.3. Combining the data envelopment analysis and the principal component analysis

All the analysis of this study was performed in the R statistical system which is free and public software environment for statistics and graphics: it is free to use, and the source code can be modified free as well (www.r-project.org). R is different from other statistical software because R is a programming language or a programming environment because it uses command scripts instead of menus. R has thousands of software packages to solve various problems (currently over 12,000). The R packages are extensions of the basic R system. For the calculations, we used the BERT tool (https://bert-toolkit.com/) that connects R and Excel, so we could work in Microsoft Excel, taking advantage of both systems' possibilities.

Using a relatively large number of input and output variables during the DEA analysis, it can cause a problem in discriminating between the efficient and the inefficient units, that is, the discrimination power decline. The Principal Component Analysis (PCA) can help both reduce the number of variables and increase the discrimination between efficient and inefficient units (Põldaru & Roots, 2014; Putri et al., 2017). The PCA is a multivariate method what can help to reduce the dimensionality of multivariate analysis (Fu & Ou, 2013; Adler & Golanyi, 2017). The PCA-DEA method is widely used in economic analysis, and it can also be seen in the portfolio analysis (Jothimani, Shankar, & Yadav, 2017). Dong, Mitchell, Knuteson, Wyman, Bussan, and Conley (2016) investigate the farm sustainability in agricultural sectors using PCA-DEA method.

We strive to minimize the loss of information using the PCA method reducing the number of model variables (Jolliffe 2002). We want to achieve this objective by transforming original variables into principal components that are linear combinations of these variables. The principal components are uncorrelated, and the first few of them contain most of the variance of the original variables (Everitt & Hothorn, 2011). PCA decomposes some correlated variables into some uncorrelated principal components using linear transformation in a multidimensional database. The extracted principal components are estimated as the projections on the eigenvectors of the covariance or correlation matrix of the database (Faed, Chang, & Saberi, 2016). The variance of the variables is an indicator describing the range of the data. The larger deviation means that the principal components include more information. The PCA ensures that the first few components will contain most of the variance of the original variables (Everitt & Hothorn, 2011). We chose the principal components which have an eigenvector value greater than 1. The remaining components can be rejected without significant loss of information. (Fu & Ou, 2013). The complete information of input and/or output variables is not lost until the principal component weights representing variables are eliminated. This procedure minimizes information loss (Andrejic et al., 2013).

The steps of the PCA-DEA analysis follow:

- Standardization of the original data (using “scale” function in R).
- Performing the principal component analysis separately for input and output variables (using “principal” function of the psych package in R).
- Determining the number of principal components (using the results of “principal” function).
- Using the principal component scores to perform DEA (using “dea” function of Benchmarking package in R)

2.4. The interval determination of DEA efficiency coefficients using Stan stochastic modeling tool

Unsal and Orkcu (2017) used simulation related to PCA-DEA analysis to increase the efficiency of company ranking. We also applied the simulation method using Stan stochastic (probabilistic) modeling tool based on Bayesian statistics. Korner-Nievergelt et al. (2015) explain that there are at least four main reasons why statistical models are used:

1. models support to describe how we think a system works,
2. data can be summarized using models,
3. comparison of model predictions with data helps to understand the system, and
4. models allow for predictions, including the quantification of their uncertainty, and, therefore, they help with making decisions.

Stan is similar to BUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) and Jags (Plummer 2003) that allow one to write a Bayesian model in a convenient language. The Bayesian approach applies the laws of probability directly to the problem which offers many advantages over the more commonly used frequentist statistical approach (Bolstad 2004). A Stan program defines a statistical model through a conditional probability function p($\theta|y;x$), where θ is a sequence of modeled unknown values, y is a sequence of modeled observed values, and x is a sequence of unmodeled predictors and constants. Stan is an imperative probabilistic programming language (Stan Development Team 2017). Stan uses Hamilton Monte Carlo (HMC) and No U-Turn Sampling (NUTS) and is implemented by C++ to speed up computation. The Stan project provides the 'rstan' package to call Stan easily from R. Hamiltonian Monte Carlo is a Markov chain Monte Carlo algorithm that avoids the random walk behavior and sensitivity to correlated parameters that plague many MCMC methods. The Hamilton Monte Carlo algorithm allows converging to high-dimensional target distributions much more quickly than simpler methods such as Gibbs sampling (Hoffman & Gelman 2012).

Box (1979) wrote, „no model is perfect, but a good model is useful". This quotation states the obvious that there is no perfect model, but if we can create such a model that reflects reality well, it can help us achieve a better understanding of the real world. The Stan modeling system can help us to build models which help us better describe the real world. The goal of the Stan project is to provide a relatively easy probabilistic programming language for Bayesian statistical modeling to fit robust, scalable, and efficient models (Carpenter et al. 2017).

Kruschke (2010) writes that applying mathematics can be very useful when the data has large variances, and we are highly uncertain as to our knowledge of the situation. Researchers very frequently encounter uncertain parameters contained in data sets with high variance, and only the statistical inference can give precise numerical bounds on our uncertainty in this situation.

3. Results

3.1. Check the number of variables and the outliers

According to Bogetoft and Otto (2011), the number of input and output variables must correspond to either of the two following expressions [(1) – (2)]whose calculated value is larger:

$$K > 3 * (m + n) \text{ or} \tag{1}$$

$$K > m * n \tag{2}$$

This requirement does not cause any problem in our case as the number of units involved in the investigation significantly exceeds both expected values. The number of units involved in the investigation (K) is 563 companies, the number of input variables (m) is 15, and the number of output variables (n) is 4. Based on the previous expressions, the expected values are: $3 * (15 + 4) = 57$ or $15 * 4 = 60$, which are considerably smaller than the number of units investigated (K = 563).

The next problem may be the number of outliers. To examine the outliers, we created boxplot-diagrams that showed that there is a large number of upper side outliers. Based on the results of the boxplot-diagrams, it can be concluded that the average number of outliers -taking into account all input and output variables – is 87, with the minimum number of 59 and the maximum number of 151. There is no significant difference between the average numbers of outliers over the years taking into account all the variables: the average number of outliers range from 84 to 91.

Subsequently, we divided the database into tithes, creating deciles. Examining the range of total data and the deciles of variables, we found that in average, 97.06% of the total range was in the tenth decile (the maximum value is 99.44% in the average of the years, and the minimum value is 87.46%). Taking into account the results, we have decided to include only those companies whose revenue values are not outliers in the last year (2014). To reduce the outliers, we have chosen the sales variable because it includes the highest outlier values based on the boxplot-diagram, although the number of outlier values is only average. The number of companies in the tenth decile is 57 in 2014, so the total number of investigated companies is 506. We can significantly reduce the outliers by this restriction, and this also meets the number of requirement for units which is 60 units.

3.2. PCA-DEA

We used input-orientated and variable return-to-scale (VRS) models in the DEA analysis. We performed the calculation of the traditional DEA efficiency coefficients for 506 companies. Both input and output

Table 2. Results of using traditional DEA

Efficiency ratio	Years						Average of years
	2009	2010	2011	2012	2013	2014	
= 0.0	466	454	458	452	475	360	444.17
< 0.1	6	3	3	2	4	4	3.67
0.1 - 0.2	0	1	0	4	3	2	1.67
0.2 - 0.3	0	2	0	3	0	5	1.67
0.3 - 0.4	1	1	2	2	0	6	2.00
0.4 - 0.5	0	3	2	4	1	7	2.83
0.5 - 0.6	1	3	0	1	1	8	2.33
0.6 - 0.7	0	1	1	3	1	10	2.67
0.7 - 0.8	0	0	1	2	0	9	2.00
0.8 - 0.9	1	0	2	0	0	15	3.00
0.9 - 1.0	0	2	1	0	0	11	2.33
= 1.0	31	36	36	33	21	69	37.67
Average efficiency	0.065	0.085	0.083	0.081	0.046	0.231	0.098

variables have been standardized. Table 2 shows that most of the companies are inefficient. In the first five years, nearly 90% of companies are inefficient, and only slightly more than 5% is efficient. The number of efficient companies has increased considerably in the last year.

Table 2 also shows the annual average of efficiency coefficients, which are very low. We can conclude that the investigated companies are inefficient considering these financial characteristics. However, the question arises whether this is in fact reality? Can this result be improved if classical DEA is combined with principal component analysis? According to the literature, the PCA-DEA combination may increase the discriminatory power of the DEA method and reduces the number of variables. The number of efficient companies in the first four years was nearly the same (≈ 6.5%), it decreased to 2/3 of the previous years in the 5th year, and it is more than doubled in the 6th year. We think that the situation cannot be quite as bad as shown in Table 2. However, we are aware that there are serious problems with the level of corporate efficiency in Hungary, and this is especially true in the case of SMEs.

Table 3 shows the results of the PCA-DEA calculation. Table 3 presents that in the case of the PCA-DEA, there were no inefficient companies, and the efficiency coefficients showed a higher scatter. We used those principal components to determine the efficiency coefficients of the DEA method whose eigenvector value had reached or exceeded 1. Table 3 also shows that instead of the 14 input variables, it was enough to use two principal components what covered approximately 90% of the total variance in the first five years. In the last year, the eigenvector value of 6 principal components was more than 1, but they only covered 82% of the total variance.

Table 3 and Table 2 show the distribution of the efficiency coefficients which significantly differ comparing the classical DEA and the PCA-DEA. We can also discover a better distribution using the PCA-DEA method. For PCA-DEA, there is a higher dispersion of company performances, that is, the efficiency coefficients of corporate performances can be more realistic. We can see

Table 3. Results of PCA-DEA calculation

Efficiency ratio	Years						Average of years
	2009	2010	2011	2012	2013	2014	
Input PCs	1, 2	1, 2	1, 2	1, 2	1, 2, 3	1, 2, 3, 4, 5, 7	
Cummulative proportion	90%	92%	92%	89%	93%	82%	
Output PCs	2, 3	1, 2	1, 2	1, 2	1, 2	1	
Cummulative proportion	76%	98%	100%	100%	100%	79%	
= 0.0	0	0	0	0	0	0	0
< 0.1	3	0	0	3	0	397	66
0.1 - 0.2	2	0	1	2	0	20	4
0.2 - 0.3	5	2	1	9	1	14	5
0.3 - 0.4	29	3	9	18	0	7	11
0.4 - 0.5	72	8	21	23	1	9	22
0.5 - 0.6	278	24	36	45	2	4	65
0.6 - 0.7	97	49	82	89	4	5	54
0.7 - 0.8	5	66	182	142	27	5	71
0.8 - 0.9	0	134	150	153	69	1	68
0.9 - 1.0	4	202	12	13	360	1	99
= 1.0	11	18	12	9	42	43	22
Average efficiency	0.554	0.836	0.738	0.710	0.933	0.147	0.653

from Table 3 that the efficiency coefficients improved significantly, and the average efficiency coefficient (0.653) is more than 6.5 times higher than in the case of traditional DEA. Average efficiency reached the highest value in the 5th year when it was close to 1 (0.933). Based on the PCA-DEA calculation it can be stated that there were no ineffective companies in this case. Figure 1 shows that the distributions of company efficiency coefficients are different, and there is only some similarity in 2010-2012 years.

3.3. Interval determination of DEA efficiency coefficients

We determined the DEA efficiency coefficients using the Stan programming language only for the year 2013. The model was executed by the 'rstan' function of the R statistical system, which translates the model into the C ++ programming language. 'RStan' is the R interface to Stan programming language. The model was run in RStudio software (Cirillo, 2016), which makes it easier to use R. Using the C ++ programming language and compiling the program allows for much faster execution. The 'rstan' program performed two thousand iterations in one Markov-chain during the simulation, one half of which was related to the prior, and the other half for the posterior estimation. The 2000-iteration simulation allows for a better estimation of the efficiency values and determines the intervals of the estimated values.

The input values of Stan model (*dea_ec*) were the efficiency coefficients obtained in the PCA-DEA method

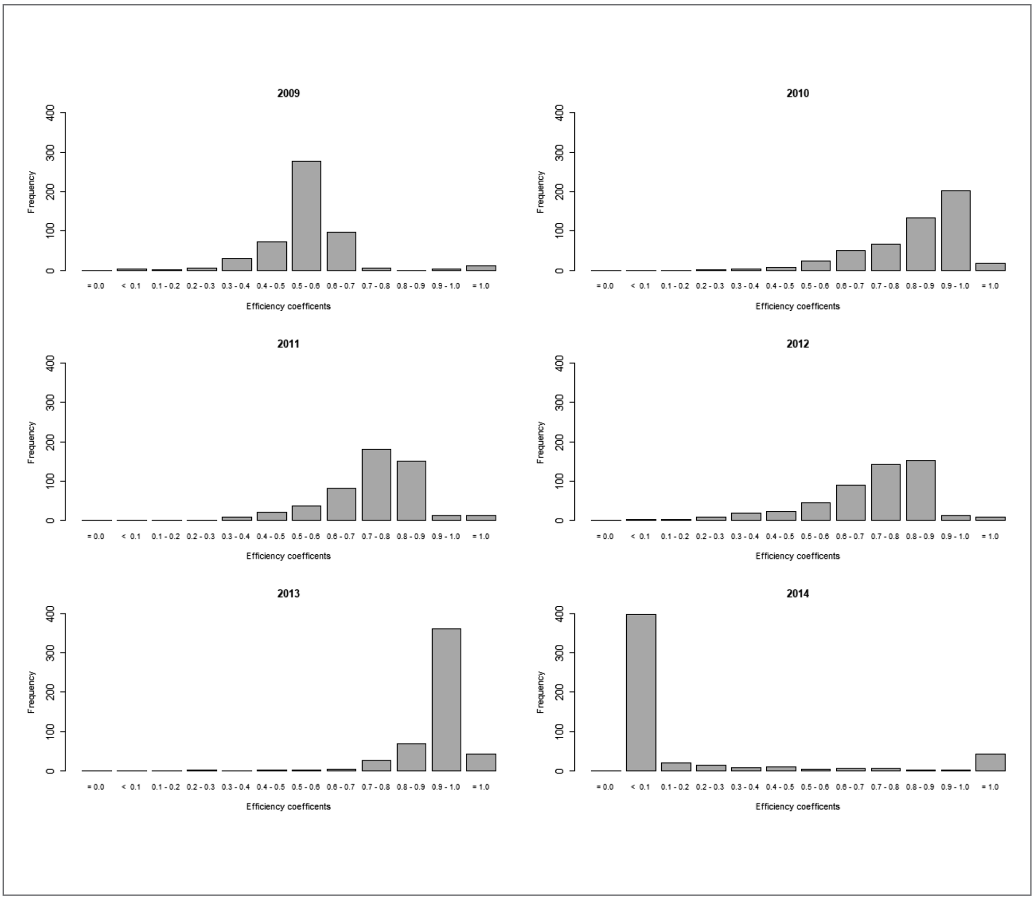

Figure 1. Results of PCA-DEA calculation

(Appendix 1). The Stan model used is necessarily a kind of extension of the DEA model.

Table 4 shows that the Stan model gave different results than the conventional DEA and the PCA-DEA method. Corporate performance indicators calculated by the Stan model have higher dispersion than observed in the previous two methods and the number of efficient companies increased.

We cannot show the intervals calculated by the Stan model in the case of all 506 companies. We present only the statistical parameters of the results of the Stan model (Table 5) in the case of all companies and efficient companies. Table 5 also shows that the average difference between the lower and the upper limits is large enough to indicate the greater uncertainty of the estimated values. If the difference between the lower and the upper limit is high, the efficiency value is more uncertain because it is in a relatively wide range; its dispersion is also higher. There are no significant differences in the proportions of the different statistical characteristics of differences considering all companies and efficient companies, but the values of efficient companies are higher (Table 5). Based on the results, we can conclude that the PCA-DEA method provided better results than the conventional DEA method, which was further improved by the Stan model. We believe that the results of the Stan model, showing that more than 28% of companies are efficient and the rest

Table 4. Results of the models

Efficiency ratio	Conventional DEA	PCA-DEA	Stan model
= 0.0	475	0	0
< 0.1	4	0	0
0.1 - 0.2	3	0	1
0.2 - 0.3	0	1	0
0.3 - 0.4	0	0	0
0.4 - 0.5	1	1	3
0.5 - 0.6	1	2	16
0.6 - 0.7	1	4	93
0.7 - 0.8	0	27	86
0.8 - 0.9	0	69	89
0.9 - 1.0	0	360	75
= 1.0	21	42	143
Average efficiency	0.046	0.933	0.851
The lower limit of average efficiency			0.827
The lower limit of average efficiency			0.850

of them are not wholly inefficient, is closer to reality than the result of the traditional DEA model.

Figure 2 also shows that while the efficiency values and their upper and lower boundaries are relatively balanced, the differences between the lower and the upper limits are dispersed quite significantly. In the case of differences, there are many outliers. There are many outliers in case of differences, which highlights the wide range between upper and lower limits.

4. Conclusion

First, this study evaluated the performance of retail food companies in Hungary's Northern Great Plain region and is based on their financial statements from 2009 to 2014 via data envelopment analysis and principal component analysis. The Conventional DEA model has been defined including 14 input and four output variables. We used various packages of the R statistical program to solve the problems.

Using the PCA-DEA method we received many different results than in the case of conventional DEA for the same companies. The PCA-DEA method evaluated the companies more extensively, while in the case of the traditional DEA the extreme values were more pronounced. The results show that the level of information reduction has a considerable effect on the efficiency classification of companies. In the case of the PCA-DEA model, the companies have relatively better distribution than in the case of conventional DEA. That is why the PCA-DEA approach is believed to produce a more accurate classification outcome.

We can conclude from the results of the Stan model that it provides us with a more acceptable result and allows a more accurate assessment of company performance. However, the results of the Stan model also draw attention to the fact that sometimes the uncertainty of the estimated values is high enough.

Table 5. Statistical characteristics of efficiency values and intervals calculated by Stan model

Name of the statistical parameter	Efficiency coefficient	Lower limit	Upper limit	Difference between upper and lower limit
		All companies		
Minimum	0.1986	0.1948	0.1982	0.0034
Quartile 1	0.7388	0.7149	0.7369	0.0220
Median	0.8803	0.8495	0.8787	0.0292
Quartile 3	1.0000	0.9663	0.9972	0.0309
Maximum	1.0000	0.9954	0.9997	0.0043
Average	0.8513	0.8266	0.8498	0.0232
		Efficient companies		
Minimum	1.0000	0.7247	0.9952	0.2705
Quartile 1	1.0000	0.9675	0.9978	0.0302
Median	1.0000	0.9833	0.9984	0.0151
Quartile 3	1.0000	0.9900	0.9988	0.0088
Maximum	1.0000	0.9954	0.9997	0.0043
Average	1.0000	0.9593	0.9982	0.0389

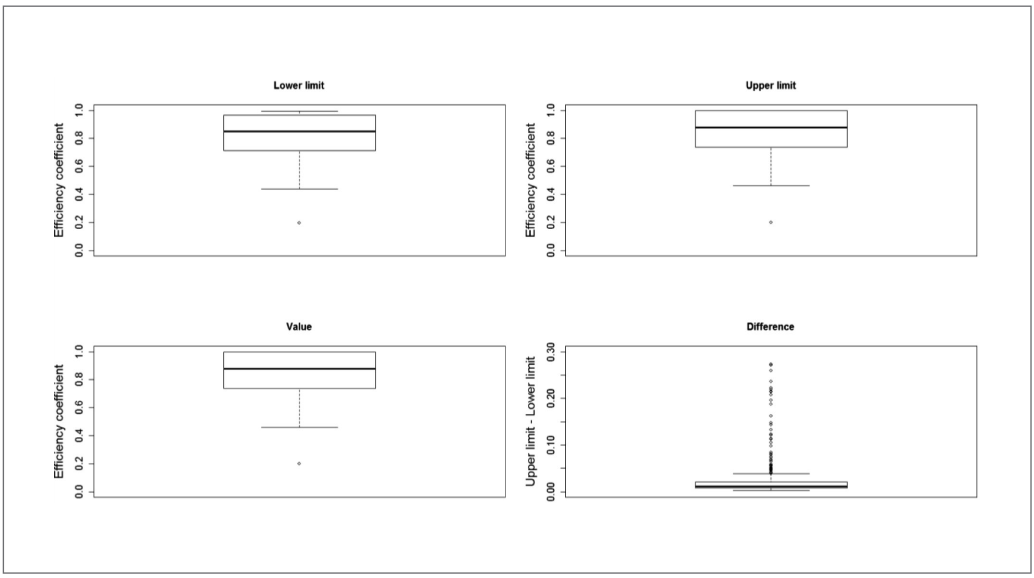

Figure 2. Box-plot diagrams of results of Stan model

In summary, it can be stated that combining different methods (DEA, PCA and Bayesian statistics) can help to define enterprise performance better. A more accurate definition of performance metrics allows for better evaluation and a well-founded comparison thus ensuring better overall decision making.

References

Adler, N., & Golany, B. (2017). PCA-DEA. In W. Cook, & J. Zhu (Eds.), Modeling Problem Structure and Data Varieties using Data Envelopment Analysis: A Problem-Solving (pp. 139-153). Boston, MA: Springer.

Andrejic, M., Bojovic, N., & Kilibarda, M. (2013). Benchmarking distribution centres using Principal Component Analysis and Data Envelopment Analysis: A case study of Serbia. Expert Systems with Applications, 40(10), 3926-3933.

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

Bolstad, W. M. (2004). Introduction to Bayesian Statistics. Hoboken, NJ: John Wiley & Sons. Inc.

Box, G.E.P. (1979, May). Robustness in the Strategy of Scientific Model Building. Defense (Technical Summary Report No. 1954). Mathematic Research Center. University of Wisconsin.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt,… Riddell, A. (2017). Stan: A probabilistic programming language. Journal of Statistical Software, 76(1), https://doi.org/10.18637/jss.v076.i01.

Charnes, A., Cooper, W. W., Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2, 429–444.

Cirillo, A. (2016). RStudio for R Statistical Computing Cookbook. Birmingham, UK: Packt Publishing

Cook, W. D., & Zhu, J. (2005). Modeling performance measurement. Applications and implementation issues in DEA. New York, NY: Springer.

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software. New York, NY: Springer.

Dong, F., Mitchell, P. D., Knuteson, D., Wyman, J., Bussan, A. J., & Conley S. (2016). Assessing sustainability and improvements in US Midwestern soybean production systems using a PCA-DEA approach. *Renewable Agriculture and Food Systems, 31*(6), 524-539

Emrouznejad, A., Parker, B. R., Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. Journal of Socio-Economic Planning Sciences, 42(3), 151–157.

Everitt. B., & Hothorn, T. (2011). An Introduction to Applied Multivariate Analysis with R. New York, NY: Springer.

Faed, A., Chang, E., & Saberi, M. (2016). Intelligent customer complaint handling utilising principal component and data envelopment analysis (PDA). Applied Soft Computing, 47, 614-630.

Farrell, M. (1957). The measurement of productive efficiency. Journal of Royal Statistical Society, Series A, 120(3), 253–281.

Fu, H. P., Ou, J. R. (2013). Combining PCA With DEA to Improve the Evaluation of Project Performance Data: A Taiwanese Bureau of Energy Case Study. Project Management Journal, 44(1), 94–106

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research 15, 1593-1623.

Huang, Z., & Li, S. X. (2011). Stochastic DEA Models With Different Types of Input-Output Disturbances. Journal of Productivity Analysis, 15, 95–113.

Jothimani, D., Shankar, R., & Yadav, S. S. (2017). A PCA-DEA framework for stock selection in Indian stock market. Journal of Modelling in Management, 12(3), 386-403

Jolliffe, I. T. (2002). Principal component analysis. New York, NY: Springer.

Korner-Nievergelt, F., Roth, T., von Felten s, Guèlat, J., Almasi, B., Korner-Nievergelt, P. (2015). Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and Stan. London, UK: Elsevier.

Kruschke, J. K. (2010). Doing Bayesian Data Analysis: A Tutorial with R and BUGS. London, UK: Elsevier.

Li, L., & Li, M. (2014), Evaluation of Energy Production Companies Efficiency Based on the Combination of Principal Component Analysis (PCA) and Data Envelopment Analysis (DEA). Environmental Engineering and Management Journal, 13(5), 1147-1154.

Lunn, D., Thomas, A., Best, N., Spiegelhalter, D. (2000). Winbugs – a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, 10, 325–337.

Plummer, M. (2003, March 20-22). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), Paper presented at the 3rd international workshop on distributed statistical computing (DSC 2003). Vienna, Austria: Technische Universität Wien. Retrieved from https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf

Põldaru, R., & Roots, J. (2014). A PCA-DEA approach to measure the quality of life in Estonian counties. Socio-Economic Planning Sciences, 48, 65-73.

Putri, E. P., Chetchotsak, D., & Jani, M. A. (2017). Performance Evaluation Using PCA and DEA: a Case Study of the Micro and Small Manufacturing Industries in Indonesia. ASR Chiang Mai University Journal of Social Sciences and Humanities, 4(1), 37-56.

Sarkar, S. (2016). Application of PCA and DEA to recognize the true expertise of a firm: a case with primary schools. Benchmarking: An International Journal, 23(3), 740-751.

Stan Development Team (2017). Stan Modeling Language: User's Guide and Reference Manual. Version 2.17.0. Available at http://mc-stan.org/users/documentation/index.html

Tavares, G. (2002, January). A Bibliography of data envelopment analysis (1978-2001) (Research Report No. 01-02). Piscataway, NJ: Rutgers Center for Operations Research, Rutgers University.

Unsal, M. G., & Orkcu, H. H. (2017), Ranking decision making units with the integration of the multidimensional scaling algorithm into PCA-DEA. Hacettepe Journal of Mathematics and Statistics, 46(6), 1187-1197.

Appendix 1.

Interval determination of the DEA coefficients in the Stan programming language

```
data {
  int<lower = 0> n;
  real<lower = 0, upper = 1> dea_ec[n];
}
parameters {
  real x1;
  real x2;
  real alpha01[n];
  real alpha02[n];
}
model {
  x1 ~ gamma(1, 1);
  x2 ~ gamma(1, 1);
  for(i in 1:n)
  {
        alpha01[i] ~ gamma(x1, x2);
        alpha02[i] ~ gamma(x1, x2);
        dea[i] ~ exponential(alpha01[i]);
        dea[i] ~ normal(0, alpha02[i]) T[0, 500];
  }
}
generated quantities {
  real ec1[n];
  real ec2[n];
  for(i in 1:n)
  {
        ec1[i] = (alpha01[i] – 1) / alpha01[i];
        ec2[i] = (alpha02[i] – 1) / alpha02[i];
  }
}
```