The International Scientific Conference INPROFORUM 2017, *November 9, 2017, České Budějovice*, 355-360, ISBN 978-80-7394-667-8.

Data Envelopment Analysis within Evaluation of the Efficiency of Firm Productivity

Miechal Houda

Abstract: *The paper deals with evaluating the efficiency of the use of production factors using methods of data envelopment analysis (DEA). We provide a first look on the problem investigated: the goal is to analyze accessible methods based on principles of data envelopment analysis and usable for evaluating the efficiency of the use of production factors, in static framework. We apply the selected method to accessible firm data from the food sector from a chosen year and perform a simple statistical inference of these results.*

Key words: Data Envelopment Analysis · Constant Returns to Scale · Productivity in the Food Industry

JEL Classification: C44 · C61

1 Introduction

Data Envelopment Analysis (DEA) is known as a non-parametric mathematical technique to evaluate relative efficiency of *decision-making units* (DMUs) based on the transformation to mathematical optimization problems. The idea of evaluating entities, for which the *efficiency* is stated in a form of the ratio

$$efficiency = \frac{output}{input}, \tag{1}$$

by a piecewise-linear convex hull of input–output pairs goes back to Farrell (1957). The idea is well developed later, especially by Charnes, Cooper, and Rhodes (1978), and Banker, Charnes, and Copper (1984); the first introduced a methodology for comparing efficiency of the units characterized by several inputs and outputs with constant returns-to-scale property, the later extended the model to variable returns-to-scale. In both cases, the efficient units are found by solving linear optimization problems, and for the inefficient ones, we are able to prescribe a way to improve inputs (or outputs) of every inefficient unit. Additional details, methods and extensions to these basic DEA models can be found in the books Cooper, Seiford, Zhu (2011), Coeli (2005), or Cooper, Seiford, Tone (2007).

To evaluate efficiency across different time points, Caves, Christensen, and Diewert (1982) introduced a productivity index as the ratio of two input distance functions, and give it a name after Malmquist (1953) exploring his concept of index numbers to evaluate price changes in the consumer context. They define the ratio as input-based Malmquist productivity index; if the overall efficiency is calculated using Farrell's idea, then the input distances are calculated as inverse of DEA efficiency scores. Färe, Grosskopf et al. (1992) extended this idea and defined the Malmquist productivity index as the geometric mean of two indexes defined in Caves, Christensen, Diewert sense. They also succeeded to decompose the index into two multiplicative components: efficiency change (the change of the efficiency of the DMU), and technical change (the shift of the technology), without additional assumptions. Many further studies extended this idea using different DEA models and under different settings to apply them in various economical areas (see e. g. Pastor, Lovell, 2005, Boussemart et al., 2009, Arabi, Munisamy, Emrouznejad. 2015 and many others).

Our intention is to analyze available data of the Czech food industry with the means of data envelopment analysis and subsequent Malmquist productivity indexes. This contribution is a very first insight into this area. We illustrate numerical issues, which are to be taken into account when developing a suitable model for the problem considered.

2 DEA Models, Returns to Scale and Production Possibility Sets

Let DMU_k denote a k-th decision making unit, $k = 1, \dots, K$ where K is the number of units in question. The input matrix is denoted $X = (x_{ik}) \in \mathbb{R}^{m \times K}$ and its rows $x_{\cdot k}$ the inputs of k-th decision making unit. Analogously, the output matrix is denoted $Y = \left(y_{jk}\right) \in \mathbb{R}^{n \times K}$ and its rows $y_{\cdot k}$ the outputs of k-th decision making unit. The unit under efficiency investigation is denoted DMU_0 through the paper.

Mgr. Michal Houda, Ph.D, University of South Bohemia in České Budějovice, Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, CZ-37005 České Budějovice, e-mail: houda@ef.jcu.cz

The central object of DEA methodology is the identification of the *production possibility sets* (PPS), in economic terms also called the *technology*. It is a set of input-output pairs, which are attainable, at least theoretically, by the production unit. For example, the simplest production possibility set is the set

$$PPS_I := \{(x, y) \mid x = X\lambda, y = Y\lambda, \lambda \in \{0,1\}^K, \sum_k \lambda_k = 1\}. \tag{2}$$

It is not hard to see that, due to the last two conditions on λ (i. e., having the binary components summing to unity), it coincides with the set $\{(x_1, y_1); \dots; (x_K, y_K)\}$ enumerating all the input-output pairs in question. Let us also characterize an efficient unit in Pareto–Koopmans dominance sense.

Definition 1. DMU_0 is *efficient* with respect to PPS if there is no other pair $(x, y) \in PPS$ dominating DMU_0, that is, having $x \leq x_{\cdot 0}$ and $y \geq y_{\cdot 0}$ with at least one (one-dimensional) inequality strict.

The set PPS_I was introduced in Bowlin et al. (1984) as the *discrete (integer) production possibility set*. The reason to choose the formulation in (2) is the associated optimization model constructed to find efficient units. Let s^- be the slack for the inequality $X\lambda \leq x_{\cdot 0}$ and s^+ the slack (surplus) for $Y\lambda \geq y_{\cdot 0}$. DMU_0 is an efficient unit if the optimal solution of the *additive (mixed integer) linear optimization problem*

$$\begin{gathered} \text{minimize } \sum_i s_i^- + \sum_j s_j^+ \text{ subject to} \\ X\lambda + s^- = x_{\cdot 0}, Y\lambda - s^+ = y_{\cdot 0}, \\ \lambda \in \{0,1\}^K, \sum_k \lambda_k = 1,\ s^-, s^+ \geq 0 \end{gathered} \tag{3}$$

has no slack greater than zero. If the unit is not found to be efficient, the only unity element of λ is pointing to the peer unit of DMU_0.

Relaxing the integer condition of (3), we define the *continuous convex production possibility set* of the form

$$PPS_C := \{(x, y) \mid x = X\lambda, y = Y\lambda,\ \lambda \geq 0,\ \sum_k \lambda_k = 1\}. \tag{4}$$

The set was introduced by Banker, Charnes, and Cooper (1984) in order to build a DEA model under *variable returns to scale*. They construct the (output oriented) BCC optimization model to verify the efficiency of DMU_0 as

$$\begin{gathered} \text{maximize } \phi + \epsilon\left(\sum_i s_i^- + \sum_j s_j^+\right) \text{ subject to} \\ X\lambda + s^- = x_{\cdot 0}, Y\lambda - s^+ = \phi y_{\cdot 0}, \\ \lambda \geq 0, \sum_k \lambda_k = 1,\ s^-, s^+ \geq 0, \phi \text{ unconstrained,} \end{gathered} \tag{5}$$

where ϵ is Archimedean infinitesimal (smaller than any positive real number). DMU_0 is found to be efficient if $\phi^* = 1$ and no slack is greater than zero. We can see from (5) that the investigated unit is compared to convex combinations (convex hull) of inputs and outputs of the units in question, exactly as PPS_C is constructed.

Removing further the convexity condition, we obtain

$$PPS_L := \{(x, y) \mid x = X\lambda, y = Y\lambda,\ \lambda \geq 0\}, \tag{6}$$

the *linear production possibility set* of Charnes, Cooper, and Rhodes (1978) corresponding to *constant returns to scale*. DMU_0 is found to be efficient solving the (output oriented) CCR optimization model

$$\begin{gathered} \text{maximize } \phi + \epsilon\left(\sum_i s_i^- + \sum_j s_j^+\right) \text{ subject to} \\ X\lambda + s^- = x_{\cdot 0}, Y\lambda - s^+ = \phi y_{\cdot 0}, \\ \lambda \geq 0,\quad s^-, s^+ \geq 0, \phi \text{ unconstrained,} \end{gathered} \tag{7}$$

under the same conditions as in the BCC case. We now compare the inputs and the outputs of DMU_0 to linear combinations (linear hull) of inputs and outputs of the units in question. The input oriented variants are constructed similarly, for example the input oriented CCR model is formulated by

$$\begin{gathered} \text{minimize } \theta - \epsilon\left(\sum_i s_i^- + \sum_j s_j^+\right) \text{ subject to} \\ X\lambda + s^- = \theta x_{\cdot 0}, Y\lambda - s^+ = y_{\cdot 0}, \\ \lambda \geq 0,\quad s^-, s^+ \geq 0, \theta \text{ unconstrained.} \end{gathered} \tag{8}$$

3 Application in Food Industry

We have considered annual accounts of 380 Czech companies from the food industry (belonging to NACE C.10 group) from the year 2014. We have chosen, for the purpose of this paper, to evaluate the companies using the input oriented model with constant returns to scale.

3.1 Grouping the companies

We provide the results for the whole C.10 group. For each company, we also have an indicator of the subgroup; the number of companies in each subgroup is given in Table 1. We have not recalculated the model for each subgroup for the purpose of this paper; rather we provide simple statistical observations based on the overall model.

Table 1 Number of companies considered according to NACE classification

C.10.1	Processing and preserving of meat and production of meat products	78
C.10.2	Processing and preserving of fish, crustaceans and molluscs	3
C.10.3	Processing and preserving of fruit and vegetables	18
C.10.4	Manufacture of vegetable and animal oils and fats	5
C.10.5	Manufacture of dairy products	28
C.10.6	Manufacture of grain mill products, starches and starch products	15
C.10.7	Manufacture of bakery and farinaceous products	109
C.10.8	Manufacture of other food products	83
C.10.9	Manufacture of prepared animal feeds	40

Source: Own processing

3.2 Considered Inputs and Outputs

As inputs of the decision-making units, we have taken these company characteristics:

- SPMAAEN: material and energy consumption. Out of 380 companies considered, 89 (23 %) reported no such costs;
- ON: personnel costs;
- STALAA: fixed assets (including buildings and equipment);
- POSN: percentage of the personnel costs.

Aside from this we have also in mind other inputs as production consumption, depreciations, tangible and intangible fixed assets, or cost of the capital.

As outputs, we have taken

- VYKONY: business performance;
- ROA: return on assets (earnings before interest and taxes per total assets). Out of 380 companies considered, 70 (18 %) have negative value of this characteristic and have to be removed from the computations.

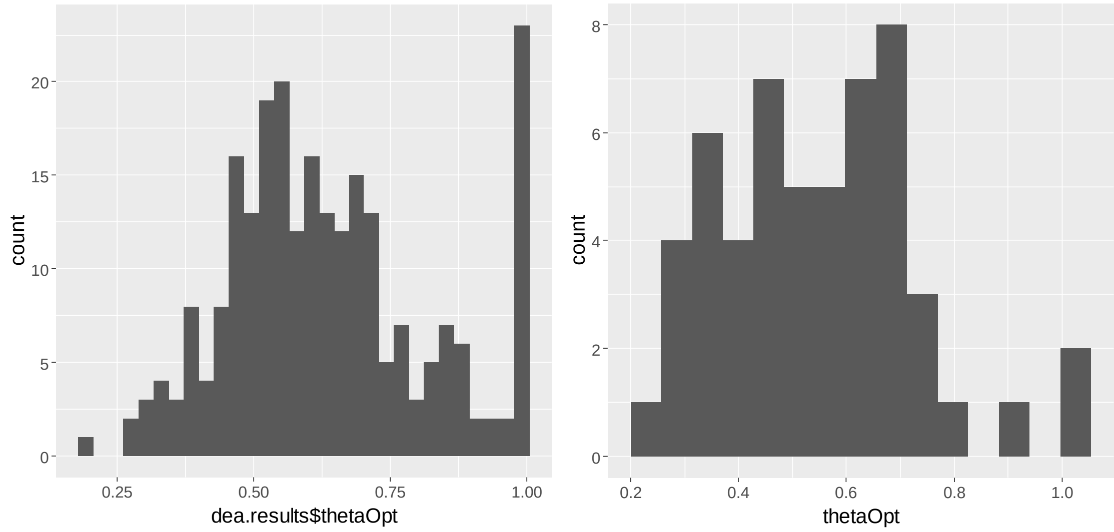
We plan to include additional outputs to the further studies, in particular the value of sales of goods and services, operating income, EBIT (earnings before interest and taxes), or value added.

In our current application, the computation carries, after removing the above-mentioned companies, 244 companies in total. Some of the removed companies exhibit both of the two mentioned issues. Another option to work with negative data is using more evolved directional DEA model, see e. g. Silva Portela, Thanassoulis, and Simpson (2004).

4 Numerical Results and Discussion

Among the 244 companies, 22 has been found efficient using the DEA model with constant returns to scale, that amounts to 9 % of evaluated (feasible) companies. Further, additional three companies exhibit the efficiency score larger than 0.95. The distribution of the efficiency scores of the companies is provided by means of histogram in Figure 1 (left). Eight of the efficient companies make part of C.10.1 group (processing and preserving of meat and production of meat products) and five of C.10.2 (manufacture of other food products). The biggest group C.10.7 (manufacture of bakery and farinaceous products) counts only two efficient companies; histogram of efficiency scores in this particular group is provided in Figure 1 (right).

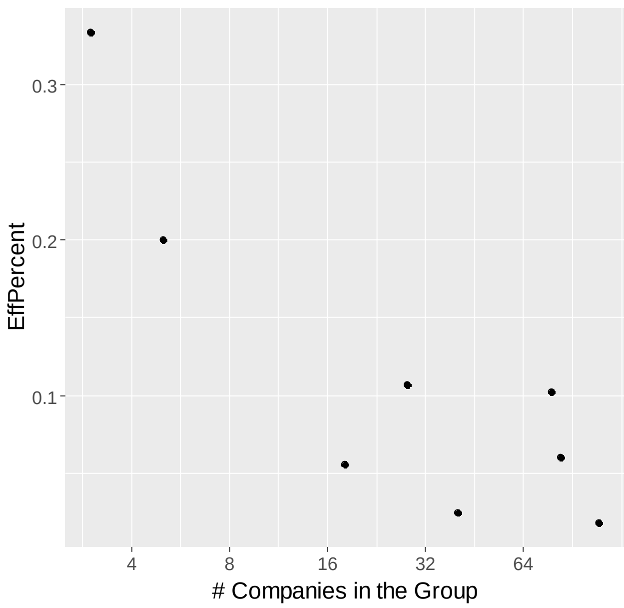
Figure 1 Histogram of efficiency scores for all companies (left), and for companies from C.10.7 group (right).



Source: Own processing.

Looking for a relationship between the number of efficient units and the size of the group, we prepared the overview given in Figure 2. The small groups (C.10.2 and C.10.4) have one efficient company each, hence resulting high percentage of efficient companies. There is no visible relationship between the group size and the number of efficient units among the other groups. (Notice also the logarithmic scale of the horizontal axis.)

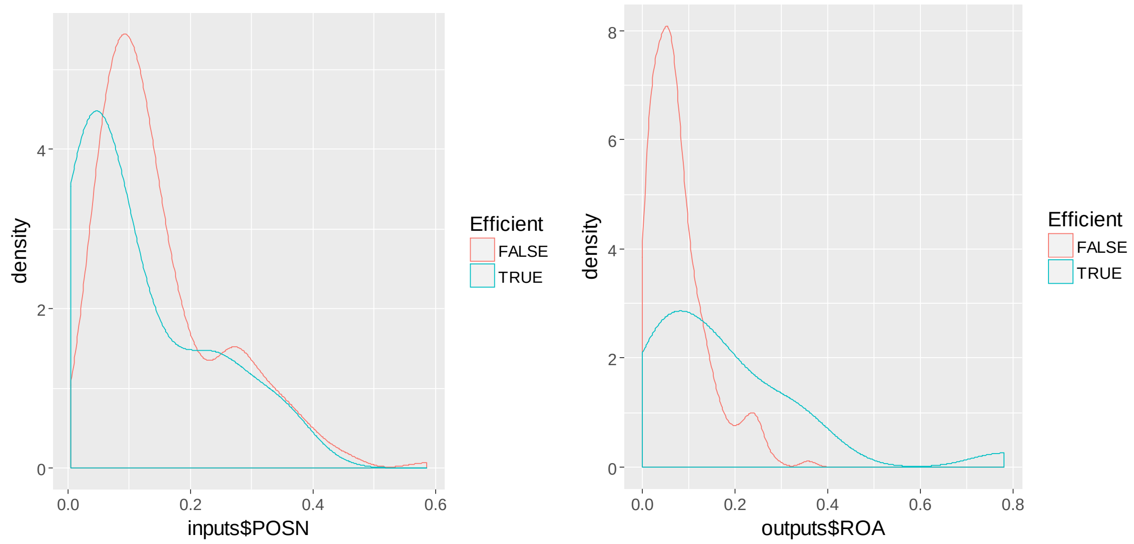
Figure 2 Relationship between the size of the group and the relative number of efficient units.



Source: Own processing.

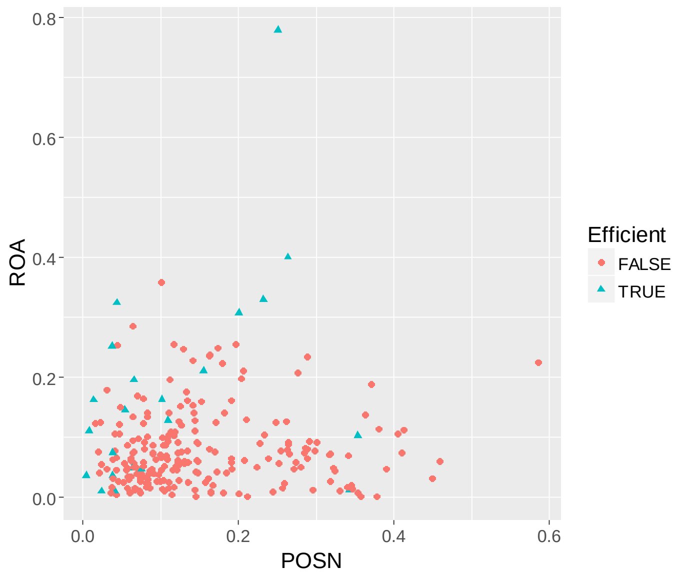
Finally, we considered a behavior of selected inputs and outputs, in particular the percentage of the personnel costs (POSN) and return on assets (ROA). The empirical distributions of these two variables with respect to the efficiency result is provided in Figure 3. The shift to the left for POSN (input variable), and to the right for ROA (output variable) is noticeable but not expressive. This can be confirmed by the scatter plot (in Figure 4) of these two variables pointing the efficiency of each unit—many units with the small output ROA are designated to be efficient while many similar units not. The similar conclusion may be made for the input variable POSN. Also notice several “extremal” efficient companies having a high input value and/or a low output value. The efficiency of these examples rely on different variables than these particular ones. Apparently, other factors included in our DEA analysis play an important role when deciding about the efficiency of decision-making units. It is necessary to be very diligent when deciding about what factors should be included into consideration.

Figure 3 Empirical distribution for values of personnel cost percentage (left) and return on assets (right).



Source: Own processing.

Figure 4 Relationship between the personnel cost percentage and return on assets



Source: Own processing

5 Conclusion

In this paper we have discussed several issues based on numerical investigation of a simple DEA model with constant returns to scale applied to the comparison of the Czech companies from the food industry. As noticed, the results are sensitive to the variable selection, which has to be done with maximal care. As more than one third of the companies was removed from the investigation due to presence of zero/negative input/outputs, it is clear that a more elaborated DEA method able to deal with such kind of the data has to be employed—the simplest CCR or BCC model will not suffice to analyze these companies. We will also consider variants of the model with different returns to scale; it is known that efficiency factors of the units is increased when moving from the constant to variable returns to scale.

References

Arabi, B., Munisamy, S., & Emrouznejad, A. (2015). A new slacks-based measure of Malmquist–Luenberger index in the presence of undesirable outputs. *Omega,* 51, 29–37.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science,* 30(9), 1078–1092.

Bowlin, W. F., Brennan, J., Cooper, W. W., & Sueyoshi, T. (1984). *DEA Models for Evaluating Efficiency Dominance*. Research Report, The University of Texas, Center for Cybernetic Studies, Austin, Texas.

Boussemart, J.-P., Briec, W., Peypoch, N., & Tavéra, C. (2009). α-returns to scale and multi-output production technologies. *European Journal of Productivity Analysis,* 197(1), 332–339.

Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica,* 50(6), 1393–1414.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research,* 2(6), 429–444.

Coelli, T. J., Prasada Rao, D. S., O'Donnell C. J., & Battes, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. 2nd ed. New York: Springer. ISBN 978-0387-24265-1.

Cooper, W. W., Seiford, L. M., & Zhu, J. eds. (2011). *Handbook on Data Envelopment Analysis*. 2nd ed. Vol. 164 of International Series in Operations Research & Management Science. New York: Springer. ISBN 978-1-4419-6150-1.

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data Envelopment Analysis. A Comprehensive Text with Models, Applications, References, and DEA-Solver Software*. 2nd ed. New York: Springer. ISBN 978-0387-45281-4.

Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1992). Productivity changes in Swedish pharmacies 1980–1989: A non-parametric Malmquist approach. *Journal of Productivity Analysis,* 3, 85–101.

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society, Series A (General),* 120(3), 253–290.

Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de Estadistica,* 4(2), 209–242.

Pastor, J. T., & Lovell, C. A. K. (2005). A global Malmquist productivity index. *Economics Letters,* 88(2), 266–271.

Silva Portella, M. C. A., Thanassoulis, E. & Simpson, G. (2004). Negative data in DEA: a directional distance approach applied to bank branches. *Journal of the Operational Research Society*, 55, 1111–1121.