National Research University Higher School of Economics

as a manuscript

Demin Sergey Stanislavovich

Data envelopment analysis methods under interval estimates and their applications

Dissertation summary

for the purpose of obtaining academic degree

Doctor of Philosophy in Applied Mathematics

Academic supervisor:
Doctor of Technical Sciences
Aleskerov Fuad Tagievich

Moscow – 2023

Problem statement

In many areas of our life we are trying to maximize the efficiency of some processes. This indicator is usually evaluated on the basis of specific features characterizing the object under study.

One of the main approaches to solve this problem is Data Envelopment Analysis. This method allows to compare the efficiency of similar objects, while the contribution of each parameter is evaluated automatically, taking into account the characteristics of each object in the form of weights for each parameter. Due to this, there is no need to do any additional analysis considering external sources for the assessment of weight coefficients for each parameter.

However, similarly to the majority of the methods, Data Envelopment Analysis requires precise data, which is not always available. Therefore, special modifications have to be used for the application, which give only approximate estimates of effectiveness.

Development of the problem

The method of Data Envelopment Analysis was initially formulated by A. Charnes, W. Cooper and E. Rhodes in 1978 based on M. Farrell's ideas about the calculation of technical efficiency of the object e_k using input (x_{jk}) and output (y_{ik}) parameters:

$$e_k = \frac{\sum_{i=1}^{M} u_i y_{ik}}{\sum_{j=1}^{N} v_j x_{jk}},$$

where M and N – numbers of output and input parameters correspondingly, u_i and v_j – nonnegative weight coefficients, which demonstrate importance of each parameter.

As a result, main optimization problem of Data Envelopment Analysis is the following (L – total number of objects in the sample):

$$\max_{u_i, v_j} \left(e_k = \frac{\sum_{i=1}^{M} u_i y_{ik}}{\sum_{j=1}^{N} v_j x_{jk}} \right)$$

$$\begin{cases} \frac{\sum_{i=1}^{M} u_i y_{il}}{\sum_{j=1}^{N} v_j x_{jl}} \leq 1 & l \in \{1, \dots, L\} \\ u_i \geq 0 & i \in \{1, \dots, M\} \\ v_j \geq 0 & j \in \{1, \dots, N\} \end{cases}$$

Later, the same authors proposed a model of Data Envelopment Analysis for the cases with variable returns on scale, which expanded the scope of the model (Banker et al., 1984):

$$\min_{\theta_k, \lambda} \theta_k$$

$$\begin{cases} -y_k + Y\lambda \geq 0 \\ \theta_k x_k - X\lambda \geq 0, \\ \lambda \geq 0 \end{cases}$$

where θ_k – efficiency of k-th object from the sample, X and Y – matrices of input and output parameters correspondingly, and λ – vector of weights showing the nearest optimal elements from the sample, which can be used as local benchmarks for the improvement of k-th object efficiency.

In the 2000s, the problem of inaccuracy in data was identified for efficiency evaluation using Data Envelopment Analysis methods. In a number of papers, it was proposed to use various methods of data preprocessing to account for possible errors in the data. For example, in (Entani et al., 2002) it was proposed to consider two scenarios and evaluate them separately – an optimistic and a pessimistic versions. In the first case, all input characteristics (used resources) of the object are assumed to be equal to the minimum values within the uncertainty intervals, and the output parameters (achieved results) are assumed to have the maximum values from these intervals. In the second case, the opposite situation occurs – the minimum values of the output parameters and the maximum values of the input parameters are used for the efficiency estimation.

An alternative approach for the problem of uncertainties in the data stems from the fuzzy set theory. For example, in (Kai & Liu, 2000) it was proposed to replace all input and output parameters with triangular fuzzy numbers in the optimization problem for the calculation of weight coefficients, while keeping the weights with real numbers. In (Lertworasirikul et al., 2003), the authors generalized this idea by replacing the decision making units features with trapezoidal fuzzy numbers.

However, the basis was still the classical model of Data Envelopment Analysis, which has its own drawbacks. As a result, at the moment there are practically no models that would take into account the lack of precise data remaining a modification of Data Envelopment Analysis and retaining positive properties such as automatic determination of the relative importance of parameters.

Problems and the aim of the study

The aim of the research is to study the properties of various models of Data Envelopment Analysis as well as to develop and subsequently test new interval models for efficiency evaluating considering inaccurate data.

Other goals of the research:

1. analyze foreign and domestic studies on the various models of Data Envelopment Analysis;
2. investigate the properties and applications of existing methods of Data Envelopment Analysis to work with inaccurate data;
3. develop new modifications of Data Envelopment Analysis considering uncertainties in the data;

4. analyze the properties of the proposed interval methods of Data Envelopment Analysis;
5. evaluate efficiency in various areas using both generated and real datasets.

Relevance of the study

The study of existing methods of Data Envelopment Analysis, development of new methods and study of their properties is very important for future application on real datasets. Firstly, this information provides a deeper understanding of the main features, advantages and disadvantages of Data Envelopment Analysis methods.

Secondly, the study of properties makes it possible to understand in which situations it is preferable to apply one or another modification to evaluate the effectiveness taking into account the specifics of an exact problem.

Thirdly, development of new methods of Data Envelopment Analysis will allow more productive consideration of available information.

Finally, Data Envelopment Analysis is applied in many different spheres. However, the results based on classic models do not always satisfy the tasks. Development of new interval models that take into account inaccuracies in the data will expand the applicability of this methodology for efficiency evaluation in all areas, thereby increasing the flexibility of this tool.

Personal contribution of the author to the research

The author developed new mathematical models. Statements on the properties of the developed interval models of Data Envelopment Analysis were also formulated and proved by the author personally. The author personally performed the software implementation of the developed Data Envelopment Analysis models and performed calculations.

In addition, in the works on evaluating the effectiveness of fire-fighting measures in the regions of the Russian Federation and evaluating the effectiveness of anti-coronavirus quarantine measures, the author personally collected and processed initial data, applied the developed interval models of Data Envelopment Analysis and analyzed the results comparing them with the classical model of Data Envelopment Analysis.

Theoretical significance of the research consists in

1. the development of new interval models of Data Envelopment Analysis;
2. study of the theoretical properties of the proposed models, which may affect the scope of application.

Practical significance of the research consists in the wide range of applications of the proposed interval models of Data Envelopment Analysis keeping the interpretability of the results. In addition the models proposed in the dissertation research can be used in similar efficiency assessment tasks regardless of the scale and technical characteristics of objects.

Main results to be defended:

1. new interval models of Data Envelopment Analysis have been developed;
2. the properties of the proposed models have been investigated;
3. the efficiency of preventive firefighting measures in different regions of the Russian Federation has been assessed;
4. the efficiency of anti-coronavirus quarantine restrictions in different countries has been evaluated.

Description of research methodology. Linear algebra methods and optimization theory methods are used to theoretically study the properties of Data Envelopment Analysis models and development of new interval modifications of

Data Envelopment Analysis. Experimental analysis involves the use of computer modeling.

Scientific novelty of the work. The following new scientific results have been obtained in the PhD thesis:

1. new interval models of Data Envelopment Analysis have been constructed using the ideas of interval scales for the parameters proposed by N. Wiener (1914, 1921);
2. the properties of the proposed models have been investigated;
3. the proposed interval models have been tested for the efficiency evaluation in different spheres and with different scales of the compared objects (countries, regions).

Summary of the work

Chapter 1 provides a general statement of the effectiveness assessment problem, recalls the history of shell data analysis and also provides an analytical review of the literature. In addition, the existing modifications of Data Envelopment Analysis are described taking into account the heterogeneity of the sample as well as the internal structure of objects.

In Chapter 2 new interval methods of Data Envelopment Analysis are proposed. At the beginning of the chapter, the models proposed in the literature for effectiveness assessment in case of inaccurate data are described. Afterwards, two new methods of Data Envelopment Analysis are proposed.

The first of them is the interval method of Data Envelopment Analysis with an optimal tube (best tube Interval DEA). According to it, at the first stage, the classical optimal efficiency boundary is constructed. Then all objects that are on the efficiency boundary as well as those that are incomparable with the optimal efficiency boundary receive a 100% efficiency value. As a result, the so-called "optimal tube" is built, which gives the name to the new method. All other objects are evaluated according to the classical approach based on the distance from the optimal boundary (Aleskerov & Demin, 2021).

The second proposed method using an interval scale of criteria is based on the idea that any parameter (both input or output) can be the most important one. Therefore, if one of the objects has the best value for at least one attribute, it should be considered as optimal. In other words, it means that the Pareto optimality principle should be taken into account.

This principle has useful properties and can be effectively applied to compare objects with interval parameter values which was investigated in (Aleskerov, 1994). According to the Pareto optimality principle the set of best objects is constructed from all objects that are not Pareto-dominated.

So, the first step of Pareto version of interval Data Envelopment Analysis (Pareto Interval DEA) is the construction of a Pareto-optimal set. All objects that fall into it automatically get 100% efficiency. For all other elements of the sample the classical model of Data Envelopment Analysis is applied (Aleskerov & Demin, 2021). Meanwhile, it is important to point out that other efficiency evaluation rules for non-optimal objects can be also used, for example, Data Envelopment Analysis with the exclusion of alternatives (Aleskerov & Petrushenko, 2015).

Section 2.3 examines the properties of the developed interval models. As a result, several statements have been proved.

Statement 1. In the case of using classic and new interval methods of Data Envelopment Analysis (best tube IDEA and Pareto IDEA) for the same data, the evaluation of the effectiveness of the object by new methods will be always not lower than the evaluation of the effectiveness by the classical method.

Proof. According to the best tube IDEA on the first step of the procedure objects lying on the best efficiency frontier $B(X)$ and objects incomparable with them are excluded from the sample with 100% efficiency. According to the Pareto

IDEA on the first step of the procedure all Pareto-optimal objects are excluded from the sample with 100% efficiency. As a result, the set of objects with 100% efficiency value might only be increased by the objects lying below classic best efficiency frontier.

In addition, the efficiency frontier for the rest of the objects might be lower because of potential adding of objects in to the group of optimal objects $B(X)$. Therefore, distance from the best efficiency frontier to all other objects cannot be higher than the distance according to the classic DEA. Consequently, their efficiency values cannot be lower than the values obtained by the classic DEA. ■

Statement 2. An increase of the uncertainty in the data parameter ε characterizing the width of the intervals (x_i^-, x_i^+) and (y_i^-, y_i^+) leads to an increase in the efficiency evaluation of objects by both interval methods of Data Envelopment Analysis.

Proof. Consider interval values of parameters as (x_i^-, x_i^+) and (y_i^-, y_i^+), which can be represented as $(x_i - \varepsilon_i, x_i + \varepsilon_i)$ and $(y_i - \delta_i, y_i + \delta_i)$ correspondingly.

Increasing of uncertainty in the data parameter ε causes increasing the width of parameter intervals ε_i and δ_i.

For instance, we consider ε_i and δ_i and increased values $\varepsilon_i{}'$ and $\delta_i{}'$.

$$\forall i,j,k \quad x_{ij} - \varepsilon_i \leq x_{ik} - \varepsilon_i \leq x_{ij} + \varepsilon_i \leq x_{ik} + \varepsilon_i \quad \Rightarrow$$
$$\Rightarrow x_{ij} - \varepsilon_i{}' \leq x_{ik} - \varepsilon_i{}' \leq x_{ij} + \varepsilon_i{}' \leq x_{ik} + \varepsilon_i{}'$$
$$\forall i,j,k \quad y_{ij} - \delta_i \leq y_{ik} - \delta_i \leq y_{ij} + \delta_i \leq y_{ik} + \delta_i \quad \Rightarrow$$
$$\Rightarrow y_{ij} - \delta_i{}' \leq y_{ik} - \delta_i{}' \leq y_{ij} + \delta_i{}' \leq y_{ik} + \delta_i{}'$$

Therefore, all pairs of intersecting intervals will be still intersecting and new pairs might occur. As a result,

$$\forall x \in X \quad x \in B(X) \Rightarrow x \in B'(X),$$

where $B(X)$ and $B'(X)$ – sets of optimal objects (Pareto optimal set for Pareto IDEA and objects inside best tube for best tube IDEA) with ε and increased ε' correspondingly. In other words, $B(X) \subseteq B'(X)$.

In addition, potential extension of $B(X)$ will narrow down $\backslash B(X)$ $\left(X\backslash B'(X) \subseteq X\backslash B(X)\right)$. As a result, efficiency evaluation of objects from $X\backslash B(X)$ might only increase because of lowering of best efficiency frontier for them. ■

Section 2.3 also describes several numerical experiments on the generated data confirming the claims about the properties of the developed interval methods of Data Envelopment Analysis. Moreover, the above-mentioned statements are preserved even in the case of certain modifications of the algorithms (for example, when taking into account the heterogeneity of the sample, as in the works of Aleskerov and Petrushchenko (2015, 2016)). Thus, it confirms that these specific features are fundamental for the proposed interval models for evaluating the functioning efficiency of similar objects.

Chapter 3 discusses the applications of the developed models. Section 3.1 examines the effectiveness of fire prevention measures in various regions of the Russian Federation. For this purpose, the regions budget expenditures on the environmental protection and the forestry as well as the number of fires in different regions have been analyzed. Based on the results of the application of various interval modifications of Data Envelopment Analysis several ratings are obtained which had minor differences in terms of the order of regions. At the same time according to interval methods the difference between the least effective regions of the Russian Federation increases, which makes it easier to identify places where it is necessary to optimize the organization of fire-fighting measures first of all.

Section 3.2 evaluates the effectiveness of anti-coronavirus quarantine measures carried out in different countries. In addition to the quarantine measures and the number of new coronavirus cases in the country, the degree of law abidingness is used to evaluate the efficiency. Moreover, the analysis of weight coefficients is carried out. As a result, it becomes possible to identify the most important types of quarantine restrictions which will allow organizing anti-coronavirus measures more effectively in future.

General conclusions of the study

Within the framework of the present dissertation research the following scientific tasks are realized:

1. new interval models of Data Envelopment Analysis have been developed;
2. the statements which allow to conclude properties and applicability of the proposed modifications of Data Envelopment Analysis have been proved;
3. the possibilities of using proposed interval models of Data Envelopment Analysis for the efficiency estimation of similar objects regardless of the scale of the objects are demonstrated. The adequacy of application of new interval models is confirmed by comparing with the efficiency estimates of classic model of Data Envelopment Analysis.

List of publications in the topic of the study

Publications from the list of journals indexed by Web of Science and Scopus international citation databases:

1. Aleskerov F., Demin S., An Assessment of the Impact of Natural and Technological Disasters Using a DEA Approach, Dynamics of Disasters

— Key Concepts, Models, Algorithms, and Insights / Ed.: P. M. Pardalos, A. Nagurney, I. S. Kotsireas, Springer, 2016, pp. 1-14.
2. Aleskerov F.T., Demin S.S., DEA for the Assessment of Regions' Ability to Cope with Disasters, Dynamics of Disasters. Impact, Risk, Resilience, and Solutions / Ed.: P. M. Pardalos, A. Nagurney, I. S. Kotsireas, A. Tsokas, Springer, 2021, Ch. 2. pp. 31-37.
3. Aleskerov F., Demin S., Myachin A., Yakuba V. Short-Term Covid-19 Incidence Prediction in Countries Using Clustering and Regression Analysis. 9th International Conference on Computers Communications and Control (ICCCC) 2022, Springer, 2023, Vol. 1435. pp. 333-342.
4. Demin S., COVID-19 Quarantine Measures Efficiency Evaluation by Best Tube Interval Data Envelopment Analysis, Operations Research Forum, Springer, 2023, 4(21).

References

1. Aleskerov F. (1994). Multicriterial Interval Choice Models. *Information Sciences*, Elsevier Inc., 80, 25-41.

2. Aleskerov F.T., Demin S.S. (2021). DEA for the Assessment of Regions' Ability to Cope with Disasters. *Dynamics of Disasters. Impact, Risk, Resilience, and Solutions*, Springer, 1(2), 31-37.

3. Aleskerov F.T., Petrushchenko V.V. (2015). An Approach to DEA for Heterogeneous Samples. Modelling, Computation and Optimization in Information Systems and Management Sciences, Advances in Intelligent Systems and Computing, Springer, 15-21.

4. Aleskerov F.T., Petrushchenko S. (2016). DEA by sequential exclusion of alternatives in heterogeneous samples. International Journal of

Information Technology and Decision Making, World Scientific Publishing Co. Pte Ltd, 15(1), 5-22.

5. Banker R.D., Charnes A., Cooper W.W. (1984). Some Models for Estimating Technical and Scale Ineffiecies in Data Envelopment Analysis. *Management Science*, INFORMS Institute for Operations Research and the Management Sciences, 30(9), 1078-1092.

6. Charnes A, Cooper WW, Rhodes E. (1978). Measuring the efficiency of decisionmaking units. *European Journal of Operations Research*, Elsevier, 2(6), 429-444.

7. Entani T., Maeda Y., Tanaka H. (2002). Dual models of interval DEA and its extension to interval data. *European Journal of Operational Research*, Elsevier, 136(1), 32-45.

8. Farrell M.J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General)*, Wiley-Blackwell Publishing Ltd, 120(3), 253-290.

9. Kao C., Liu S.T. (2000). Fuzzy efficiency measures in data envelopment analysis. *Fuzzy Sets and Systems*, Ed.: I. Couso, B. De Baets, L. Godo, Elsevier, 113(3), 427-437.

10. Lertworasirikul S., Fang S.-C., Joines J.A., Nuttle H.L.W. (2003). Fuzzy data envelopment analysis (DEA): a possibility approach. *Fuzzy Sets and Systems*, Ed.: I. Couso, B. De Baets, L. Godo, Elsevier, 139(2), 379-394.

11. Wiener N. (1914). A contribution to the theory of relative position. *Proceedings of the Cambridge Philosophical Society*, Cambridge Philosophical Society, 17, 441-449.

12. Wiener N. (1921). A new theory of measurement: a study in the logic of mathematics. *Proceedings of the London Mathematical Society*, John Wiley and Sons Ltd, 19, 181-205.