ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS

Volume 67 137 Number 6, 2019

DATA ENVELOPMENT ANALYSIS MODELS IN NON-HOMOGENEOUS ENVIRONMENT

Josef Jablonský[1]

[1] Department of Econometrics, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 13067 Praha 3, Czech Republic

To link to this article: https://doi.org/10.11118/actaun201967061535
Received: 3. 6. 2019, Accepted: 9. 9. 2019

To cite this article: JABLONSKÝ JOSEF. 2019. Data Envelopment Analysis Models in Non-Homogeneous Environment. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis,* 67(6): 1535–1540.

Abstract

Data envelopment analysis (DEA) is a non-parametric method that is widely used for relative efficiency and performance evaluation of the set of decision-making units (DMUs). It is based on maximization of a weighted sum of outputs produced by the unit under evaluation divided by the weighted sum of inputs of the same unit, and the assumption that this ratio for all other units has to be lower or equal to 1. An important assumption for applications of DEA models is the homogeneity of the units. Unfortunately, the homogeneity assumption is not fulfilled in many real applications. The paper deals with the analysis of efficiency using DEA models in the non-homogeneous environment. One of the problems lies in non-homogeneous outputs. In this case, the units under evaluation spend the same inputs but produce completely or at least partly different set of outputs. The paper formulates several models how to deal with this problem and compares the results on a numerical example. Other main sources of non-homogeneity are discussed as an excellent possible starting point for future research.

Keywords: data envelopment analysis, non-homogeneous units, missing data, efficiency, performance

INTRODUCTION

Data envelopment analysis (DEA) is a tool for measuring the relative efficiency and performance of a set of decision-making units (DMUs). It is based on the comparison of the weighted sum of outputs produced the DMU under evaluation and the weighted sum of its inputs needed for the production of outputs. Let us consider the set of DMUs containing n elements. The DMUs are described by m inputs and r outputs with input and output values x_{ij}, $i = 1, \ldots, n$; $j = 1, \ldots, m$, and y_{ik}, $i = 1, \ldots, n$; $k = 1, \ldots, r$, respectively. The efficiency score of the DMU$_q$ (denoted as θ_q) is defined as the weighted sum of outputs divided by the weighted sum of inputs with the weights v_i, $i = 1, \ldots, m$ and u_k, $k = 1, \ldots, r$:

$$\theta_q = \frac{\sum_{k=1}^{r} u_k y_{qk}}{\sum_{j=1}^{m} v_j x_{qj}}. \qquad (1)$$

Traditional DEA model introduced by Charnes *et al.* (1978) assumes homogeneity of the set of DMUs. Their model is formulated as follows:

$$\text{Maximize} \qquad \theta_q = \frac{\sum_{k=1}^{r} u_k y_{qk}}{\sum_{j=1}^{m} v_j x_{qj}}, \qquad (1)$$

$$\text{subject to} \qquad \frac{\sum_{k=1}^{r} u_k y_{ik}}{\sum_{j=1}^{m} v_j x_{ij}} \le 1, \qquad i = 1, \ldots, n,$$

$$v_j \ge \varepsilon, \qquad j = 1, \ldots, m,$$

$$u_k \ge \varepsilon, \qquad k = 1, \ldots, r.$$

This model is not linear in its objective function but can be moved to a linear program using Charnes-Cooper transformation easily. Depending on the way of transformation the results is either input-oriented or output-oriented version of the model. The input-oriented formulation follows directly from model (1):

$$\text{Maximize} \qquad \theta_q = \sum_{k=1}^{r} u_k y_{qk},$$

$$\text{subject to} \qquad \sum_{k=1}^{r} u_k y_{ik} - \sum_{j=1}^{m} v_j x_{ij} \le 0, \; i = 1, \ldots, n,$$

$$\sum_{j=1}^{m} v_j x_{qj} = 1, \tag{2}$$

$$v_j \ge \varepsilon, \qquad j = 1, \ldots, m,$$

$$u_k \ge \varepsilon, \qquad k = 1, \ldots, r.$$

Models (1) and (2) are often denoted as CCR models. They assume constant returns to scale for the construction of the envelope of the data set. They were generalized and extended by Banker *et al.* (1984) for the assumption of variable returns to scale. They are denoted as BCC models. Both groups of models belong among the main DEA models applied in economic analyses.

Unfortunately, the assumption of homogeneity of all units of the set is too restrictive and often does not correspond to reality. Therefore, a research effort in this field is concentrated on the development of models that relax the assumption of perfect homogeneity, especially in the last years. The survey in Material and methods section of the paper presents the most essential published results dealing with non-homogeneity in the data set in applications of DEA models. A problem of missing data in outputs is considered and possible ways how to deal with this problem are proposed in section Results. The same section contains the results of numerical experiments and the final section concludes the paper and discusses future research.

MATERIALS AND METHODS

Non-homogeneity in DEA models is one of the research topics in efficiency and performance evaluation. The following survey is the selection of several papers published in the last years.

Dyson *et al.* (2001) was probably the first paper that pointed out the problems with the homogeneity assumptions in DEA models. The purpose of this paper was to highlight some of the pitfalls that have been identified in DEA applications and to suggest protocols to avoid the pitfalls and guide the application of the methodology. Castelli *et al.* (2001, 2004) considered the problem of evaluation of a set of interdependent decision making subunits that make up larger decision-making units. In general, the subunits do not meet homogeneity assumptions, and they need not spend the same set of inputs and/or produce the same set of outputs. Haas and Murphy (2003) presented three adjustment techniques in order to overcome heterogeneity based on traditional DEA models. Saen *et al.* (2005) presented a methodology for dealing with DEA models with missing data that is based on the combination of traditional DEA methodology and analytic hierarchy process. The papers Cook *et al.* (2012, 2013) summarized the sources of non-homogeneity and dealt with a related problem of missing or imprecise data in outputs in DEA. The study Imanirad *et al.* (2013) extended the traditional DEA methodology to allow efficiency measurement in situations where only partial input-to-output impacts exist. The paper Li *et al.* (2016) considered efficiency where DMUs are non-homogeneous on the input side and examines the case where different DMUs have different natural resource configurations. Zha *et al.* (2013) studied the problem of dealing with missing data in applications of DEA models. Cross-efficiency evaluation is one of the approaches for ranking of efficient units in DEA models. Zhu *et al.* (2017) extends this concept for non-homogeneous DMUs.

Network systems with serial, parallel or combined structure are usually non-homogeneous. Efficiency analysis in this kind of models is of high importance. Du *et al.* (2015) and Barat *et al.* (2018) focused on heterogeneity in network DEA models. The paper Singh and Ranjan (2017) is focused on efficiency analysis of non-homogeneous parallel systems for the performance measurement in higher education which is an important topic not only in our conditions.

Among the application-oriented papers, Huang *et al.* (2016) presents a study on the evaluation of hotels in Taiwan with non-homogeneity in hotel types where the hotels are divided into two non-homogeneous groups. Li *et al.* (2018) is an application paper that deals with efficiency evaluation of non-homogeneous Hong Kong hospitals. Sun *et al.* (2017) presents an application of heterogeneous DEA models in performance evaluation of bank systems. Chen *et al.* (2018) proposes an original procedure for efficiency evaluation of two non-homogeneous groups and illustrates its properties on an example from the academic environment.

RESULTS

The sources of non-homogeneity in the data set can be of a very different nature. In this paper, we deal with non-homogeneity in outputs which belongs to one of the most common possibilities. Let us suppose that the set of DMUs is divided into 2 disjoint groups. The DMUs from both groups spend the same inputs, but the outputs are wholly or partly different. This case will be illustrated in a simple example.

The example is modified from its original source Chen *et al*. (2018). It is not a real case study, but it is used rather for illustrative purposes. Let us suppose 23 departments in total. All of them spent two inputs for their activities – total expenses in thousands of local currency (X1) and the total number of full-time employees (X2). They are divided into two groups – first of them consists of 14 departments that are teaching and research oriented. They are characterized by two outputs – the number of accreditation certificates (Y1) and the number of journal publications (Y2). The second group of departments (9 units) is teaching oriented only, and they are described just by one output that is the same as the first output of the first groups of departments. The complete data set is included in Tab. I.

Let us suppose there are no other data available for the allocation of funds among the departments for the next planning period. That is why it is necessary not to evaluate the two groups of departments independently but to compare all 23 units. The results of this evaluation can serve as the initial tool for allocation of funds among departments. There are proposed several approaches how to deal with missing data when the DMUs are divided into two or more generally into several disjoint groups. We will further compare the results of three of them:

1. The missing values are replaced by zeros, and a traditional DEA model is applied in order to derive efficiency scores of all units.
2. The missing values are replaced by pessimistic, optimistic, or the most likely values, and again a traditional DEA model is applied.
   In our example, the number of journal publications per one researcher (employee) varies from 0.481 for Department 7 until 2.301 for Department 12 with a mean value of 1.052. Using these coefficients three scenarios are created, and the missing values are replaced by multiplying the number of employees by three coefficients for each Department.
3. This approach was proposed in Cook *et al.* (2012). It is based on a reduction of inputs by 100·α % for the first group of units (set D_1) where α is a constant that determines the part of inputs needed for the output(s) that are not produced by the second group of departments (set D_2).
   Let us suppose that 100·α % of the capacity of the departments in D_1 is used for research, i.e. for a "production" of journal papers. The process of deriving the final efficiency scores for all departments consists of two steps:
   - In the first step, we use 100·(1 - α) % of both inputs and the outputs produced by both groups of units. This calculation is performed for all units of the set. The efficiency scores given in this step for the units in D_2 are taken as their final scores.
   - The second step considers the units in D1 only. This step uses 100·α % of all inputs and the outputs that are not produced by the units in D_2 (the number of journal papers in our case). The final efficiency scores for the units in D_1 are the weighted averages of their scores derived in both steps.

Tab. II presents the results given by all three approaches. In all calculations, we have applied traditional CCR model (2). In order to differentiate among efficient units, many approaches have been proposed in the past. Andersen and Petersen (1993) proposed their super-efficiency model that is used in our numerical experiments. While traditional model (2) assigns efficiency scores 1 for efficient units, super-efficiency model relaxes this limit, and originally efficient units have their super-efficiency scores greater (or equal) than 1. Tab. II contains efficiency scores of all departments calculated by model (2) and its super-efficiency modification under the following assumptions:

I: *Data set for 23 University departments*

	X1	X2	Y1	Y2
Department 1	54284	1934	91	1794
Department 2	48879	454	26	744
Department 3	96830	1269	49	800
Department 4	54595	790	22	470
Department 5	85779	1391	21	1281
Department 6	46512	1122	39	1481
Department 7	53834	1798	90	865
Department 8	70780	536	25	774
Department 9	72830	845	20	1287
Department 10	53207	1501	91	1624
Department 11	90563	487	50	585
Department 12	53271	641	85	1475
Department 13	47361	1685	31	1722
Department 14	70816	861	77	1201
Department 15	51629	872	25	
Department 16	23814	240	28	
Department 17	21179	609	22	
Department 18	32474	732	27	
Department 19	47789	974	52	
Department 20	53584	263	28	
Department 21	49254	223	33	
Department 22	30295	682	46	
Department 23	34424	259	58	

Source: Chen *et al.* (2018)

- Missing values are replaced by zeros (column M1 in Tab. II).
- Missing values are replaced by pessimistic, average, and optimistic values as discussed above (columns M2 – pessimistic, M3 – most likely, M4 – optimistic).
- The model presented in Cook *et al.* (2012) is applied. Final efficiency scores for the departments in D_1 are computed as a simple average of the scores derived using two steps as described above, i.e. there are used identical weights for both steps (column M5).

All calculations were performed using own original procedures written in LINGO modelling system which is an ideal tool for solving DEA models because makes it possible to simplify notation of all models and their modifications.

DISCUSSION

The results presented in Tab. II show a very close similarity of efficiency and super-efficiency scores computed by M1 model (missing values replaced by zeros) and M2 pessimistic model. The efficiency scores of departments in D_2 tend to increase in case of average and optimistic model (M3 and M4). Optimistic model M4 leads to maximum efficiency of all departments in D_2 which is an unacceptable conclusion. The results obtained by the last model (M5) always show lower efficiency (super-efficiency) scores than model M1 to M3, but the correlation between the pairs of efficiency scores given by models M1 to M3 and M5 are very high (always higher than 0.9). On the contrary,

III: *Correlation coefficients*

	M1	M2	M3	M4	M5
M1	1.0000	0.9991	0.9416	0.3006	0.9520
M2		1.0000	0.9445	0.3155	0.9558
M3			1.0000	0.5225	0.8856
M4				1.0000	0.3122
M5					1.0000

Source: own processing

II: *Results – efficiency scores*

	M1	M2	M3	M4	M5
Department 1	1.0585	1.0585	1.0585	0.9944	0.9446
Department 2	0.7122	0.7122	0.7122	0.7122	0.5542
Department 3	0.3142	0.3140	0.3137	0.3072	0.3049
Department 4	0.3012	0.3012	0.3012	0.2585	0.2755
Department 5	0.5109	0.5109	0.5109	0.4002	0.3308
Department 6	0.9907	0.9907	0.9907	0.5855	0.7447
Department 7	0.9775	0.9775	0.9775	0.9775	0.7147
Department 8	0.6275	0.6275	0.6275	0.6275	0.4629
Department 9	0.6619	0.6619	0.6619	0.6619	0.4192
Department 10	1.0316	1.0309	1.0299	1.0225	0.9708
Department 11	0.6714	0.6568	0.6332	0.5220	0.5885
Department 12	1.6180	1.6180	1.6180	1.0000	1.3727
Department 13	1.1003	1.1003	1.1003	0.5753	0.7415
Department 14	0.6761	0.6753	0.6742	0.6590	0.6463
Department 15	0.2854	0.2960	0.6025	1.0000	0.2079
Department 16	0.6965	0.6966	0.7080	1.0000	0.5816
Department 17	0.6074	0.6074	0.9119	1.2757	0.4252
Department 18	0.4881	0.4881	0.7511	1.0000	0.3485
Department 19	0.6398	0.6398	0.7161	1.0000	0.4603
Department 20	0.4754	0.5265	0.6095	1.0000	0.4754
Department 21	0.6608	0.6938	0.7475	1.0000	0.6608
Department 22	0.8914	0.8914	0.8984	1.1218	0.6365
Department 23	1.6208	1.6208	1.6208	1.6208	1.1821

Source: own processing

the optimistic model leads to completely different results that are hardly acceptable. Correlation coefficients between the pairs of efficiency scores obtained by all models M1 to M5 are presented in Tab. III.

We can conclude by this that all models except M4 can be considered as a starting point for discussion about resource allocation among all departments. This task is widely discussed not only in DEA literature and its importance for practice is clear without any doubts.

CONCLUSION

Analysis of efficiency in case on non-homogeneity of data sets is an important task because the perfect homogeneity is often difficult to reach. This paper dealt with only one particular case of non-homogeneity. We analyzed the case of two groups of units and missing outputs for one of them. This approach can be generalized for more groups and different structure of missing values. The contribution of the paper consists in a proposal of a new procedure how to deal with missing data in DEA models. The results show that this procedure is simple and leads easily to acceptable conclusions that can serve as a starting point for further analyses, e.g. resource allocation among DMUs.

Non-homogeneity in DEA models can be given by many other factors than the one case that was analyzed in this paper. The research in this area is open, and this paper is just a starting point for a broader research under the research project mentioned below.

Acknowledgements

The research is supported by the Grant Agency of the Czech Republic, project No. 19-08985S – *Models for efficiency and performance evaluation in non-homogeneous economic environment*.

REFERENCES

ANDERSEN, P. and PETERSEN, N. C. 1993. A procedure for ranking efficient units in data envelopment analysis. *Management Science*, 39(10): 1261–1264.

BANKER, R. D., CHARNES, A. and COOPER, W. W. 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9): 1078–1092.

BARAT, M., TOHIDI, G. and SANEI, M. 2018. DEA for nonhomogeneous mixed networks. *Asia Pacific Management Review*, 24(2): 161–166.

CASTELLI, L., PESENTI, R. and UKOVICH, W. 2001. DEA-like models for efficiency evaluations of specialized and interdependent units. *European Journal of Operational Research*, 132(2): 274–286.

CASTELLI, L., PESENTI, R. and UKOVICH, W. 2004. DEA-like models for the efficiency evaluation of hierarchically structured units. *European Journal of Operational Research*, 154(2): 465–476.

CHARNES, A., COOPER, W. W. and RHODES, E. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6): 429–444.

CHEN, Y., WU, L. and LU, B. 2018. Data envelopment analysis procedure with two non-homogeneous DMU groups. *Journal of Systems Engineering and Electronics*, 29(4): 780–788.

COOK, W. D., HARRISON, J., ROUSE, P. and ZHU, J. 2012. Relative Efficiency Measurement: The Problem of a Missing Output in a Subset of Decision Making Units. *European Journal of Operational Research*, 220(1): 79–84.

COOK, W. D., HARRISON, J., IMANIRAD, R., ROUSE, P. and ZHU, J. 2013. Data envelopment analysis with nonhomogeneous DMUs. *Operations Research*, 61(3): 666–676.

DU, J., CHEN, Y. and HUO, J. 2015. DEA for non-homogenous parallel networks. *Omega*, 56(1): 122–132.

DYSON, R. G., ALLEN, R., CAMANHO, A. S. PODINOVSKI, V., SARRICO, C. S. and SHALE, E. A. 2001. Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132(2): 245–259.

HAAS, D. and MURPHY, F. 2003. Compensating for non-homogeneity in decision-making units in data envelopment analysis. *European Journal of Operational Research*, 144(3): 530–544.

HUANG, C. W., CHIU, Y. H., TU, C. H., LUO, Z. Y. and WANG, Z. 2016. Using the non-homogeneous frontier two-stage DEA model to assess the efficiencies of expense utilization and operation of the Taiwanese hotel industry. *International Transactions in Operational Research*, 23(6): 1067–1087.

IMANIRAD, R., COOK, W. D. and ZHU, J. 2013. Partial input to output impacts in DEA: Production considerations and resource sharing among business subunits. *Naval Research Logistics*, 60(3): 190–207.

KAO, C. and HWANG, S. N. 2008. Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. *European Journal of Operational Research*, 185(1): 418–429.
LI, W. H., LIANG, L., COOK, W. D. and ZHU, J. 2016. DEA models for non-homogeneous DMUs with different input configurations. *European Journal of Operational Research*, 254(3): 946–956.
LI, Y., LEI, X. and MORTON, A. 2018. Performance evaluation of nonhomogeneous hospitals: The case of Hong Kong hospitals. *Health Care Management Science*, 22(2): 215–228.
SAEN, R. F., MEMARIANI, A. and LOTFI, F. H. 2005. Determining relative efficiency of slightly non-homogeneous decision making units by data envelopment analysis: a case study in IROST. *Applied Mathematics and Computation*, 165(2): 313–328.
SINGH, S. and RANJAN, P. 2018. Efficiency analysis of non-homogeneous parallel sub-unit systems for the performance measurement of higher education. *Annals of Operations Research*, 269(1–2): 641–666.
SUN, J., WANG, C., JI, X. and WU, J. 2017. Performance evaluation of heterogeneous bank supply chain systems from the perspective of measurement and decomposition. *Computers & Industrial Engineering*, 113: 891–903.
ZHA, Y., SONG, A., XU, C. and YANG, H. 2013. Dealing with missing data based on data envelopment analysis and halo effect. *Applied Mathematical Modelling*, 37(9): 6135–6145.
ZHU, W., YU, Y. and SUN, P. 2018. Data envelopment analysis cross-like efficiency model for non-homogeneous decision-making units: The case of United States companies' low-carbon investment to attain corporate sustainability. *European Journal of Operational Research*, 269(1): 99–110.

Contact information

Josef Jablonský: jablon@vse.cz