DOI: 10.7763/IPEDR. 2013. V59. 7

Data Envelopment Analysis Network Models with Interval Data

Josef Jablonsky[+]

University of Economics, Prague, W. Churchill Sq. 4, 130 67 Praha 3, Czech Republic

Abstract. Data envelopment analysis (DEA) is a non-parametric method for relative efficiency evaluation of decision making units described by multiple inputs and multiple outputs. It is based on solving linear programming problems. Since 1978 when basic DEA model was introduced many its modifications were formulated. Two or, in more general, multi-stage models with series or parallel structure (network models) belong among them. Standard DEA models are based on deterministic inputs and outputs. The paper deals with DEA network models under the assumption that inputs and/or outputs are continuous interval variables. Under this assumption the efficiency scores of production units are random variables as well. Several approaches for description of random efficiency scores were developed for standard DEA models but only few for models with network structure. They are mostly based on formulation of linear optimization problems. Another methodological approach for DEA models with interval data is simulation. The paper compares results given by simulation experiments and by optimization DEA network models with interval data.

Keywords: Data Envelopment Analysis, Network Models, Interval Data, Efficiency.

1. Introduction

Data envelopment analysis (DEA) is a non-parametric technique for evaluation of relative efficiency of decision making units described by multiple inputs and outputs. Let us suppose that the set of decision making units (DMUs) contains n elements. The DMUs are evaluated by m inputs and r outputs with input and output values x_{ij}, $i = 1, 2,\dots, m$, $j = 1,2,\dots, n$ and y_{kj}, $k = 1, 2,\dots,r$, $j = 1, 2,\dots, n$, respectively. The efficiency of the DMU$_q$ can be expressed as the weighted sum of outputs divided by the weighted sum of inputs with weights reflecting the importance of single inputs/outputs. The efficiency score of the DMU$_q$ can be given by solving the optimization problem (1) which is standard CCR model formulated by Charnes et al. in [2].

Minimize θ_q

Subject to (1)

$$\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta_q x_{iq}, \quad i = 1,2,\dots,m,$$

$$\sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{kq}, \quad k = 1,2,\dots,r,$$

$$\lambda_j \geq 0, \quad j = 1,2,\dots,n,$$

Where $\lambda_j, j = 1, 2,\dots, n$ are weights of DMUs, s^-_i, $i = 1, 2,\dots, m$, and s^+_k, $k = 1, 2,\dots, r$ are slack (surplus) variables and θ_q is the efficiency score of the DMU$_q$ which expresses necessary reduction of inputs in order this unit becomes efficient. If the optimal value of model (1) $\theta^*_q = 1$, then the DMU$_q$ is CCR efficient and it is lying on the CCR efficient frontier, otherwise the unit is not CCR efficient. Model (1) is CCR model with input orientation, i.e. this model looks for reduction of inputs in order to reach the efficient frontier. The output oriented modification of the presented model is straightforward. The BCC model under variable returns to scale assumptions originally presented in [1] extends the formulation (1) by convexity constraint $\sum_j\lambda_j = 1$. Both mentioned DEA models measure efficiency of a transformation of m inputs into r outputs in

[+] Corresponding author. Tel.: + (420) 224095403; fax: + (420) 224095423.
E-mail address: jablon@vse.cz

one stage and under an assumption that all data are deterministic but the production process is often more complex and the data may be stochastic. The paper formulates two-stage network model with serial structure where outputs of the first stage are inputs of the second stage and offers its solution under assumption that all inputs and outputs are interval variables.

The paper is organized as follows. The next section contains basic formulation of DEA network model with interval data and discusses possibility of its solution using optimization and simulation models. Section 3 presents results of numerical experiments based on real data set – evaluation of efficiency of bank branches. Final part of the paper summarizes presented results and discusses directions for future research.

2. DEA Network Models with Interval Data

Model (1) measures the relative efficiency of one-stage transformation of m inputs into r outputs. The transformation of inputs into final outputs can be considered as a two- or even several-stage process. The inputs of the first stage are transformed into its outputs and all or at least some of these outputs are used as inputs of the second stage that produces final outputs. Let us denote the input values of the first stage x_{ij}, $i = 1, 2,\ldots, m$, $j = 1, 2,\ldots, n$ and the output values of the first stage y_{ij}, $i = 1, 2,\ldots, r$, $j = 1, 2,\ldots, n$. Supposing that all outputs of the first stage are used as inputs of the second stage and that the final output values are z_{ij}, $i = 1, 2,\ldots, p$, $j = 1, 2,\ldots, n$. Two-stage DEA models are widely analyzed and discussed within professional community. Theoretical issues can be found e.g. in [6]. Among numerous case studies can be mentioned papers [5] and [7]. Two-stage DEA model under constant returns to scale assumption can be formulated according to [3] as follows:

Minimize $\theta_q - \phi_q$

Subject to

$$\sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_q x_{iq}, \qquad i = 1,2,\ldots,m,$$

$$\sum_{j=1}^{n} y_{kj}\lambda_j \ge \tilde{y}_{kq}, \qquad k = 1,2,\ldots,r,$$

$$\sum_{j=1}^{n} y_{kj}\mu_j \le \tilde{y}_{kq}, \qquad k = 1,2,\ldots,r, \tag{2}$$

$$\sum_{j=1}^{n} z_{lj}\mu_j \ge \phi_q z_{lq}, \qquad l = 1,2,\ldots,p,$$

$$\theta_q \le 1, \phi_q \ge 1,$$

$$\lambda_j \ge 0, \mu_j \ge 0, \qquad j = 1,2,\ldots,n,$$

Where λ_j and μ_j, $j = 1, 2,\ldots, n$, are weights of the DMUs in the first and second stage, θ_q and φ_q are efficiency scores of the DMU$_q$ in the first and second stage and $\tilde{y}_{kq}$ are variables to be determined. The DMU$_q$ is recognized as efficient by model (2) if the efficiency scores in both stages are $\theta_q = 1$ and $\varphi_q = 1$ respectively, and the optimal objective value of the presented model is 0. The inefficient units can be ranked relatively by the following geometric average efficiency measure:

$$e_q = (\theta_q / \varphi_q)^{1/2} \tag{3}$$

Inputs and outputs in both stages usually reflect past values of the DMUs and the model (2) supposes that the inputs and outputs of the units are given as deterministic values. For evaluation and estimation of future efficiency of the DMUs it can be useful to consider inputs and outputs as random variables. They can be given as interval values or more generally as random variables with defined continuous probabilistic distribution. Approaches for dealing with interval data in DEA can be divided into two groups – optimization and simulation methods. Optimization approaches are based on solving one or several linear programs and result to an index or indices for each DMU that can be used for their ranking. One of the optimization approaches is presented in [4]. It supposes that the corresponding values for inputs and outputs are continuous variables with uniform distribution defined over intervals $x_{ij} \in \langle x^L_{ij}, x^U_{ij} \rangle$, $y_{kj} \in \langle y^L_{kj}, y^U_{kj} \rangle$ and $z_{lj} \in \langle z^L_{lj}, z^U_{lj} \rangle$. Efficiency scores of the DMUs under the assumption of interval inputs and outputs are random variables defined over an interval. The lower bound for efficiency score in both stages is given by taking into account the worse inputs and outputs for the evaluated unit DMU_q and the best characteristics for all the other units and similarly the upper bound is defined by using the best characteristics of the evaluated unit and the worse ones for all the other units. The optimization model for deriving the lower bound for efficiency score of the unit DMU_q in the first stage is as follows:

Minimize θ_q

Subject to

$$\sum_{j=1, j\neq q}^{n} \left(\lambda_j x_{ij}^L\right) + \lambda_q x_{iq}^U + s_i^- = \theta x_{iq}^U, \quad i = 1,2,\ldots,m,$$

$$\sum_{j=1, j\neq q}^{n} \left(\lambda_j y_{kj}^U\right) + \lambda_q y_{kq}^L - s_k^+ = y_{kq}^L, \quad k = 1,2,\ldots,r, \tag{4}$$

$$\lambda_j \geq 0,\ s_k^+ \geq 0,\ s_i^- \geq 0.$$

According to the values of the lower and upper bounds of efficiency scores in each of two stages the DMUs can be divided into three subsets E^1, E^2 and E^3:

E^1: DMUs always efficient – this subset contains units that are efficient in any case, i.e. even their inputs and outputs are on their worst values and the inputs and outputs of other units are on their best bounds.

E^2: DMUs conditionally efficient by suitable adjusting of inputs and outputs of all or at least some of the units (upper bound of their efficiency score is 1).

E^3: DMUs never efficient (upper bound of their efficiency score is lower than 1).

This approach can lead to quite different results, e.g. a DMU can belong to the set E^1 (always efficient) in the first stage and to the set E^2 or even E^3 (never efficient) in the second stage. In order to evaluate the efficiency of both stages simultaneously model (2) can be modified according to model (4) - the modified optimization model considers interval variables of the inputs of the first stage and final outputs and intermediate characteristics are taken into account on their average level. The results of the modified model (2) allow dividing the units into three classes as presented above.

Except optimisation models simple simulation tools can be used to analyse the presented problem even under a more general assumptions than interval values of random variables. Simulation approach is more time consuming than the optimisation one but it gives much more information that can be useful for a detailed analysis of the problem. This approach can be simply described by the following steps:

- Random generation of all random variables of the model
- Modified two-stage DEA model (2) is solved with the values generated in the previous step
- Information from the random trials are processed and evaluated by means of a suitable software tool

Simulation trials offer much more information about distribution of efficiency scores of particular units comparing to above described optimization procedure. Results given by both - optimization and simulation – procedure are compared on a simple example in the next section.

3. Numerical Example

Applications of DEA models are numerous. Results of the above formulated models will be illustrated on an example of 67 selling branches of one of the Czech mobile phone operators. The example takes into account two inputs (operational expenses in CZK and the number of business hours per year), two intermediate characteristics (the number of transactions of current customers and the number of new customers), and two final outputs (financial contribution of the branch in CZK and the ICCA score measuring the customer satisfaction). The data for all 67 branches are available with a certain level of uncertainty. That is why the fixed data from 2010 are used for numerical experiments together with modified set of data. In this modification we suppose that the data are independent continuous random variables with uniform distribution over interval <0.95*x*, 1,10*x*>, where *x* is the original fixed value. Computational experiments are divided into two groups.

Table 1: Results of efficiency evaluation (two single stages)

DMU	1st stage		2nd stage		Geometric average	Rank
	lower	upper	lower	upper		
1	0.9014	1.0000	0.3382	0.6128	0.6556	9
2	1.0000	1.0000	1.0000	1.0000	1.0000	1
3	0.7139	0.9889	0.6986	1.0000	0.7847	4
4	0.8948	1.0000	0.6082	1.0000	0.8496	2
5	0.8572	1.0000	0.5228	0.9757	0.8038	3
6	0.6191	1.0000	0.4905	0.8589	0.6650	8
7	0.5876	1.0000	0.5133	0.9546	0.6801	7
8	0.5662	0.8217	0.9951	1.0000	0.6931	6
9	0.7996	1.0000	0.3609	1.0000	0.7422	5
10	0.5745	0.9452	0.4094	0.7141	0.5695	10

(1) Efficiency evaluation of the branches in two separated stages using model (5) and its modifications under variable returns to scale assumption. Lower and upper bounds for efficiency scores of the branches in both stages are main results of applied models. The branches can be ranked according to several criteria in this case – e.g. geometric average of maximum or minimum efficiencies in both stages. The results of this approach for a selection of 10 branches are presented in Table 1. According to the results of the first stage there is only one unit (DMU_2) that is always efficient, 6 units are conditionally efficient and the remaining ones are always inefficient. Similar conclusions hold for the second stage – again the unit DMU_2 is always efficient, other 4 units are conditionally efficient and 5 units are inefficient even they work on their best bound. The pre-last column of Table 1 contains geometric average of two values – simple average of lower and upper values in each stage. The last column presents ranking of the selected units according to the geometric average in the pre-last column. It is really questionable what criterion to use for ranking of units when both stages are separated. It is clear that the most efficient unit is the unit DMU_2 that is always efficient in both stages. Next places in ranking are occupied by units that are conditionally efficient at least in one of two stages.

(2) Evaluation of efficiency using modified model (4), i.e. considering both stages simultaneously. Optimization approach leads to lower and upper bounds for efficiency scores as above. Simulation approach was realized with uniformly generated data of all units and optimization run with model (4). After 50 trials some information about distribution of efficiency scores in both stages and overall efficiency are given.

Table 2 contains information about lower and upper bounds of efficiency scores given by modified model (4) - θ_q and φ_q values are synthesized by (5). Next three columns present similar information from 50 simulation trials as described in previous section of the paper – minimum, maximum and average efficiency characteristics (5). Last two columns of Table 2 compare rankings of DMUs by the middle of interval given by lower and upper bounds from optimization runs on the one side and by average characteristics (5) from simulation approach on the other side. It is clear that both rankings are very close each other. Of course simulation procedure offers much more information to decision makers than single optimization approach. Information about distribution of efficiency scores is not presented here due to a limited space for the paper. Comparison of results given by two single models for efficiency evaluation (Table 1) and the two-stage model (Table 2) shows more significant differences in final ranking of DMUs – e.g. DMU_9 is on 5^{th} place in two single models approach and it is one of the worse DMUs when two-stage model is applied.

Table 2: Results of efficiency evaluation (two-stage model)

DMU	Optimization		Simulation (50 trials)			Rank optim	Rank simulation
	lower	upper	lower	upper	avg		
1	0.6138	0.7233	0.6174	0.7212	0.6635	8	7
2	1.0000	1.0000	1.0000	1.0000	1.0000	1	1
3	0.6491	0.8865	0.6814	0.8302	0.7507	4	5
4	0.7873	0.9537	0.8067	0.9487	0.8629	2	2
5	0.6326	0.8982	0.6908	0.8906	0.7813	5	4
6	0.5826	0.7564	0.6155	0.6994	0.6514	7	8
7	0.6295	0.8317	0.6832	0.7646	0.7240	6	6
8	0.7417	0.8796	0.7683	0.8444	0.8164	3	3
9	0.4936	0.8246	0.4955	0.8124	0.5865	9	10
10	0.5803	0.7201	0.5947	0.6770	0.6324	10	9

4. Conclusions

Efficiency evaluation of network production systems is a very complex task. The paper is focused on a simplest system which is two-stage serial model. Under the assumption of deterministic data there are formulated several DEA models for efficiency evaluation. In case of stochastic data one can use optimization approach that offer information about the worse and the best efficiencies (under worse and best conditions for the evaluated unit) only. The same information can be given by simulation approach but except results given by optimization approach many other results can be of an interest for decision makers. Both approaches are illustrated on a simple numerical example of a real-world nature. Further research will be focused on more complex network systems with serial or parallel structures.

5. Acknowledgements

The research is supported by the project of the Czech Science Foundation P403/12/1387.

6. References

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

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

[3] Y. Chen, L. Liang and J. Zhu. Equivalence in two-stage DEA approaches. *European Journal of Operational Research*, 2009, **193** (3): 600-604.

[4] D. Despotis and Y. Smirlis. Data envelopment analysis with imprecise data. *European Journal of Operational Research*, 2002, **140** (1): 24-36.

[5] J. Jablonsky. A two-stage AHP/DEA model for evaluation of efficiency. In: T. Subrt (ed.), *Mathematical Methods in Economics 2009*. Praha: University of Life Sciences Press. 2009, pp. 159 - 164.

[6] L. Liang, Z. Q. Li, W. D. Cook and J. Zhu. Data envelopment analysis efficiency in two-stage networks with feedback. *IIE Transactions*, 2011, **43** (5): 309-322.

[7] J. C. Paradi, C. Rouatt and J. Zhu. Two-stage evaluation of bank branch efficiency using data envelopment analysis. 2011, *Omega*, **39** (1): 99-109.