Two-Stage Data Envelopment Analysis Model with Interval Inputs and Outputs

Josef Jablonsky

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-stage 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 decision making 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.**

Index Terms—**Data envelopment analysis, efficiency, interval data, two-stage model.**

I. Introduction

Data envelopment analysis (DEA) is a non-parametric technique for evaluation of relative efficiency of decision making units characterized 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}, $i = 1, 2,\ldots, m$, $j = 1, 2,\ldots, n$ and y_{kj}, $k = 1, 2,\ldots, r$, $j = 1, 2,\ldots, n$, respectively. The efficiency of the q-th DMU can be expressed as the weighted sum of outputs divided by the weighted sum of outputs with weights that reflect the importance of single inputs v_i, $i = 1, 2,\ldots, m$, and outputs u_k, $k = 1, 2,\ldots, r$ as follows:

$$\theta_q = \frac{\sum_{k=1}^{r} u_k y_{kq}}{\sum_{i=1}^{m} v_i x_{iq}}.$$

Standard CCR input oriented DEA model formulated by Charnes et al. [1] consists in maximization of efficiency score of the DMU_q subject to constraints that efficiency scores of all other DMUs are lower or equal than 1. The linearized form of this model is as follows:

Maximize

$$\theta_q = \sum_{k=1}^{r} u_k y_{kq}$$

subject to (1)

$$\begin{aligned} &\sum_{i=1}^{m} v_i x_{iq} = 1, \\ &\sum_{k=1}^{r} u_k y_{kj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, && j = 1,2,\ldots,n, \\ &u_k, v_i \ge \varepsilon, && k = 1,2,\ldots,r, i = 1,2,\ldots,m. \end{aligned}$$

If the optimal value of the 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. The model (1) is often referenced as primal CCR model. Its dual form is more convenient from the computational point of view and its mathematical model is as follows:

Minimize

$$\theta_q$$

subject to (2)

$$\begin{aligned} &\sum_{j=1}^{n} x_{ij} \lambda_j + s_i^- = \theta_q x_{iq}, && i = 1,2,\ldots,m, \\ &\sum_{j=1}^{n} y_{kj} \lambda_j - s_k^+ = y_{kq}, && k = 1,2,\ldots,r, \\ &\lambda_j \ge 0, \quad j = 1,2,\ldots,n, \end{aligned}$$

where λ_j, $j = 1, 2,\ldots, n$ are weights of DMUs, s^-_i, $i = 1, 2,\ldots, m$, and s^+_k, $k = 1, 2,\ldots, r$ 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 Banker et al. [2] extends the formulation (1) by convexity constraint $\sum_j \lambda_j = 1$. More information about DEA models and about their numerous modifications including MS Excel solver for DEA models is included in Zhu [3].

DEA models are applied very often in many fields of human activities. They find numerous applications in

Manuscript received January 12, 2013; revised February 27, 2013. The research is supported by the project of the Czech Science Foundation P403/12/1387.

Josef Jablonsky is with the Department of Econometrics, University of Economics, Prague, 13067 Czech Republic (e-mail: jablon@ vse.cz).

evaluation of efficiency in public sector (education, health care, regional studies, etc.), finance (evaluation of bank and insurance companies' branches), industries (branches of public or private companies), and many other fields. Several applications in the Czech Republic are described e.g. in Dlouhy et al. [4], and Grmanova and Jablonsky [5].

The above presented introductory DEA models measure efficiency of the transformation of *m* inputs into *r* outputs in one stage and under the 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 can be taken as inputs of the second stage and offers its solution under the assumption that all inputs and outputs are random variables.

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

II. Two-Stage DEA Model with Interval Data

Models (1) and (2) measure 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 multi-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. The two-stage production process can be expressed as it is on Fig. 1.

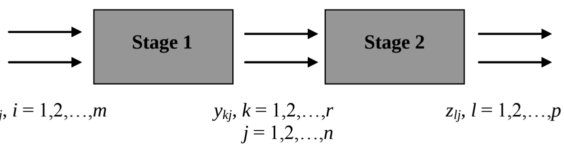

Fig. 1. Two-stage production process.

Let us denote the input values of the first stage x_{ij}, $i = 1, 2,\dots, m$, $j = 1, 2,\dots, n$ and the output values of the first stage y_{kj}, $k = 1, 2,\dots, r$, $j = 1, 2,\dots, n$. Let us suppose that all outputs of the first stage are used as inputs of the second stage and that the final output values are z_{lj}, $l = 1, 2,\dots, p$, $j = 1, 2,\dots, n$. Two-stage DEA models are widely analyzed and discussed within professional community. Theoretical issues can be found e.g. in Liang et al. [6]. Among numerous case studies the papers Jablonsky [7] and Paradi *et. al*. [8] can be mentioned. Two-stage DEA model under constant returns to scale assumption can be formulated according to Chen et al. [9] as follows:

Minimize

$$\theta_q - \phi_q$$

subject to (3)

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

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

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

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

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

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

where λ_j and μ_j, $j = 1,2,\dots,n$, are weights of the DMUs in the first and second stage, respectively, 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 (3) 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:

$$e_q = (\theta_q / \varphi_q)^{1/2} \qquad (4)$$

Inputs and outputs in both stages usually reflect past values of the DMUs. That is why the model (3) 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 random data in DEA can be divided into two groups – optimization and simulation. 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 Despotis and Smirlis [10]. 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 using of 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 of the production process is as follows:

Minimize

$$\theta_q$$

subject to (5)

$$\sum_{j=1,\, j\ne q}^{n} \left(\lambda_j x^L_{ij}\right) + \lambda_q x^U_{iq} + s_i^- = \theta x^U_{iq}, \qquad i = 1,2,\dots,m,$$

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

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

The appropriate model for calculation of upper bound of the efficiency score of the unit DMU_q is as follows:

Minimize

$$\theta_q$$

subject to (6)

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

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

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

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 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 using model (4) the optimization model must consider random variables of first stage inputs' and final outputs and intermediate characteristics use in their average level. The model (3) is then modified according to the models (5) and (6), i.e. using lower/upper bounds for characteristics in the first and last group of constraints. The results of the modified model (3) allow dividing of the units into three classes as above: always efficient (efficient in both stages), never efficient (inefficient in both stages), and conditionally efficient.

One of the disadvantages of the optimisation approach for dealing with random inputs and outputs in DEA models consists in the necessity to consider just interval values (uniform distribution). Except optimisation models simple simulation tools can be used to analyse the presented problem under a more general assumptions. 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:

1) Generation of all random variables of the model. This step can be simply realized within MS Excel environment by means of built-in functions or VBA procedures.
2) Modified two-stage DEA models (3) are solved with the values generated in the previous step. In our experiments the LP solver included in the LINGO modelling system was used which is powerful enough and allows a simple linking with MS Excel sheets.
3) Information from the random trials are processed and evaluated by means of suitable software tools. A MS Excel add-in application for Monte Carlo experiments (e.g. Crystal Ball or @RISK) is a possible alternative.

Simulation trials give 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.

III. Computational Experiments

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 model for efficiency analysis is presented on Fig. 2. The following inputs, intermediate characteristics (outputs of the first stage and inputs of the second one) and final outputs are taken into account:

Inputs:

- Operational expenses (rental costs, wages and overheads), and
- Number of business hours per year is an important characteristic influencing total number of transactions (one of the outputs of the first stage).

Intermediate characteristics:

- Number of transactions of current customers, and
- Number of transactions of new customers.

Outputs:

- Financial contribution of the branch in CZK (Czech crowns).

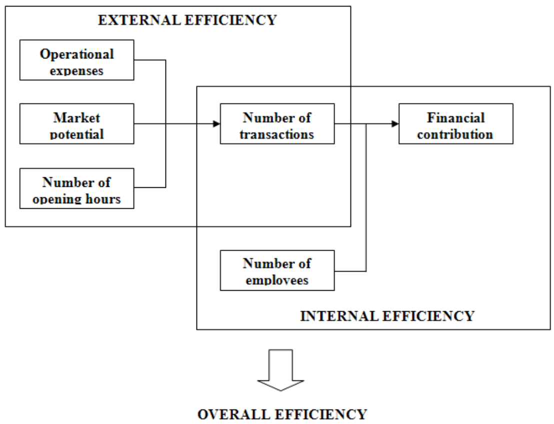

Fig. 2. Model for efficiency analysis of selling branches.

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 phases.

1) The first phase is evaluation of efficiency of the branches in two separated stages using the models (5) and (6) under

variable returns to scale assumption. Lower and upper bounds for efficiency scores of the branches in both stages are the main results of applied models. The branches can be ranked according to several criteria – e.g. geometric average of maximum or minimum efficiencies in both stages. The lower and upper bounds for efficiency scores given by the models (5) and (6) for a selection of 10 branches are presented in Table I. 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 bounds.

TABLE I: RESULTS FOR TWO SINGLE STAGES

DMU	1st stage		2nd stage	
	lower	upper	lower	upper
1	0.9014	1.0000	0.3382	0.6128
2	1.0000	1.0000	1.0000	1.0000
3	0.7139	0.9889	0.6986	1.0000
4	0.8948	1.0000	0.6082	1.0000
5	0.8572	1.0000	0.5228	0.9757
6	0.6191	1.0000	0.4905	0.8589
7	0.5876	1.0000	0.5133	0.9546
8	0.5662	0.8217	0.9951	1.0000
9	0.7996	1.0000	0.3609	1.0000
10	0.5745	0.9452	0.4094	0.7141

2) Evaluation of efficiency using modified model (3), 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 (3). After 50 trials some information about distribution of efficiency scores in both stages and overall efficiency are given. Table II contains information about lower and upper bounds of efficiency scores given by modified model (3) - θ_q and φ_q values are synthesized using formula (4). Next three columns present similar information from 50 simulation trials as described in previous section of the paper – minimum, maximum and average efficiency scores calculated using the application of models (5) and (6).

TABLE II: RESULTS FOR TWO- STAGE MODEL

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

The comparison of results given by two presented approaches is included in Table III. The first two columns of this table contain ranking of DMUs by efficiency scores calculated by means of the models (5) and (6) in two single stages. The column "*Geom*" contains geometric average of two values – simple average of lower and upper bounds of the first stage (first two columns of Table I) and simple average of lower and upper bounds of the second stage (last two columns of Table I). Ranking of DMUs according to this criterion is presented in the next column of Table III. In is really questionable what criterion could be used for ranking of DMUs when both stages are taken as independent. It is clear that the most efficient unit is the unit DMU_2 that is always efficient in both stages. Next rankings are occupied by units that are conditionally efficient at least in one of the two stages.

TABLE III: COMPARISON OF RESULTS

DMU	Two single stages		Two-stage model			
	Geom.	Rank	Optim	Rank	Simul	Rank
1	0.6556	9	0.6686	8	0.6635	7
2	1.0000	1	1.0000	1	1.0000	1
3	0.7847	4	0.7678	4	0.7507	5
4	0.8496	2	0.8705	2	0.8629	2
5	0.8038	3	0.7654	5	0.7813	4
6	0.6650	8	0.6695	7	0.6514	8
7	0.6801	7	0.7306	6	0.7240	6
8	0.6931	6	0.8107	3	0.8164	3
9	0.7422	5	0.6591	9	0.5865	10
10	0.5695	10	0.6502	10	0.6324	9

The remaining columns of Table III contain rankings of DMUs by the middle of the interval given by lower and upper bounds from optimization runs on the one side and by average characteristics (4) from simulation approach on the other side. It is clear that both rankings are very close each other but, of course, the simulation procedure offers much more information to decision makers than single optimization approach. Very interesting may be the information about the distribution of efficiency scores. This paper does not contain any more detailed information about this distribution due to the limited space of the paper.

Comparison of results given by two single models for evaluation of efficiency and the two-stage model presented in Table III shows more significant differences in final ranking of DMUs – e.g. DMU_9 is ranked as fifth unit in two single models approach and it is one of the worse DMUs when two-stage model is applied. That is why it is of a high importance to study network models instead to analyze the efficiency independently in two single stages.

IV. CONCLUSIONS

Evaluation of efficiency of network production systems is a very complex task. The paper is focused on a simplest system which is two-stage serial DEA 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 this many other results can be of 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.

REFERENCES

[1] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, June 1978.

[2] R. D. Banker, A. Charnes, and W. W. Cooper, “Some models for estimating technical and scale inefficiencies in data envelopment analysis,” *Management Science*, vol. 30, no. 9, pp. 1078–1092, September 1984.

[3] J. Zhu, *Quantitative Models for Performance Evaluation and Benchmarking*: *Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*, Boston: Kluwer Academic Publishers, 2002.

[4] M. Dlouhy, J. Jablonsky, and I. Novosadova, “Using data envelopment analysis for efficiency evaluation of Czech hospitals,” *Politicka ekonomie*, no. 1, pp. 60-71, February 2007.

[5] E. Grmanova and J. Jablonsky, “Efficiency analysis of Slovak and Czech insurance companies using data envelopment analysis models,” *Ekonomicky casopis*, vol. 57, no. 9, pp. 857-869, September 2009.

[6] L. Liang, Z. Q. Li, W. D. Cook, and J. Zhu, “Data envelopment analysis efficiency in two-stage networks with feedback,” *IIE Transactions*, vol. 43, no. 5, pp. 309-322, May 2011.

[7] J. Jablonsky, “A two-stage AHP/DEA model for evaluation of efficiency,” in *Proceedings of the International Conference Mathematical Methods in Economics 2009*, ed. T. Subrt, pp. 159-164, Praha, University of Life Sciences Press. 2009.

[8] J. C. Paradi, C. Rouatt, and J. Zhu, “Two-stage evaluation of bank branch efficiency using data envelopment analysis,” *Omega*, vol. 39, no. 1, pp. 99-109, January2011.

[9] Y. Chen, L. Liang, and J. Zhu, “Equivalence in two-stage DEA approaches,” *European Journal of Operational Research*, vol. 193, no. 3, pp. 600-604, April 2009.

[10] D. Despotis and Y. Smirlis, “Data envelopment analysis with imprecise data,” *European Journal of Operational Research*, vol. 140, no. 1, pp. 24-36, January 2002.

Josef Jablonsky was born in the city of Rychnov nad Kneznou, Czech Republic in 1955. He received his master degree in Mathematical Methods in Economics at the University of Ecoomics, Prague, Czech Republic in 1978. He earned his Ph.D. degree in the same field and the same university in 1983. In 2003 he was appointed as the full professor and the Head of Department of Econometrics at the University of Economics, Prague.

Since 1983 he is a member of the Department of Econometrics at the University of Economics, Prague, Czech Republic. He published several books and a large number of articles in journals and conference proceedings. His book *Operational Research* is the most widespread textbook in this field in the Czech Republic. His professional interests include multiple criteria decision making, analytic hierarchy and network process, data envelopment analysis, mathematical modeling and optimization, and software for mathematical modeling. In 2007 he was the main organizer of EURO XXII Conference (the Annual Conference of the Association of the European Operational Research Societies) with more than 2000 participants.