Data envelopment analysis models with network structure

Josef Jablonský[1]

Abstract. Data envelopment analysis (DEA) is a non-parametric technique for evaluation of relative efficiency of decision making units described by multiple inputs and outputs. It is based on solving linear programming problems. Since 1978 when basic DEA model was introduced many its modifications were formulated. Among them are two or multi-stage models with serial or parallel structure often called network DEA models that are widely discussed in professional community in the last years. The paper presents several approaches for analysis of network production systems and formulates DEA network models for two-stage serial and parallel systems. The presented models are used in a case study that is focused on evaluation of production and profit efficiency of banks in the Czech Republic. The results given by two single DEA models for the first and second stages and then by network model are presented and compared.

Keywords: data envelopment analysis, two-stage models, network models, efficiency

JEL Classification: C44
AMS Classification: 90C15

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}, $i = 1,2,\ldots,m$, $j = 1,2,\ldots,n$ 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 reflecting the importance of single inputs/outputs v_i, $i = 1,2,\ldots,m$ and u_k, $k = 1,2,\ldots,r$:

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

Standard CCR input oriented DEA model formulated in [3] Charnes et al. consists in maximization of efficiency score (1) 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:

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

subject to (2)

$$\sum_{i=1}^{m} v_i x_{iq} = 1,$$

$$\sum_{k=1}^{r} u_k y_{kj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1,2,\ldots,n,$$

$$u_k, v_i \ge \varepsilon, \quad k = 1,2,\ldots,r, i = 1,2,\ldots,m.$$

If the optimal value of the model (2) $\theta^*_q = 1$ 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 (2) is often referenced as primal CCR model. Its dual form is common and its mathematical model is as follows:

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 4, 13067 Praha 3, Czech Repubic, e-mail: jablon@vse.cz

Minimze θ_q

subject to (3)

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

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

$$\lambda_j \geq 0, j = 1,2,\ldots,n,$$

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.

The models (2) and (3) are CCR models with input orientation, i.e. they look for reduction of inputs in order to reach the efficient frontier. The output oriented modification of the presented models is straightforward. The BCC models under variable returns to scale assumptions originally presented in [2] by Banker et al. extend the formulation (3) by convexity constraint $\sum_j\lambda_j = 1$. The presented basic DEA models measure efficiency of a transformation of m inputs into r outputs in one stage and under an assumption that all data are deterministic but the production process is often much more complex. That is why various modifications of standard DEA models were formulated in the last years. One stream in this research is represented by network DEA models. They suppose that production process contains several stages with serial or parallel structures.

The paper deals with parallel network models, formulates some of the network models and illustrates their application for two-stage efficiency evaluation of Czech banks. The paper is organized as follows. The next section contains basic formulation of DEA network models. Section 3 presents a simple two-stage model for efficiency evaluation of Czech banks and discusses results of numerical experiments. Final part of the paper summarizes presented results and discusses directions for future research.

2 Network DEA models

The models (2) and (3) measure the relative efficiency of one-stage transformation of m inputs into r outputs. The transformation of inputs into final outputs can be taken as a two- or several-stage process. The inputs of the first stage are transformed into its outputs and all or at least some of these outputs are utilized as inputs of the second stage that are using for production of 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_{kj}, $k = 1,2,\ldots,r$, $j = 1,2,\ldots,n$. Supposing that all outputs of the first stage are taken as inputs of the second stage and that the final output values are z_{lj}, $l = 1,2,\ldots,p$, $j = 1,2,\ldots,n$. Two-stage DEA models are widely analyzed and discussed within professional community. Theoretical issues can be found e.g. in [7]. Among numerous case studies can be mentioned papers [5] and [8]. One of the first two-stage serial DEA models is formulated in [6]. Below is its dual formulation with constant returns to scale assumption and input orientation:

Minimize θ_q

subject to

$$\sum_{j=1}^{n} x_{ij}\lambda_j \leq \theta_q x_{iq}, \quad i = 1,2,...,m,$$

$$\sum_{j=1}^{n} y_{kj}\lambda_j - \sum_{j=1}^{n} y_{kj}\mu_j \geq 0, \quad k = 1,2,...,r, \tag{5}$$

$$\sum_{j=1}^{n} z_{lj}\mu_j \geq z_{lq}, \quad l = 1,2,..., p,$$

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

where λ_j and μ_j, $j = 1,2,\ldots,n$, are weights of the DMUs in the first and second stage, and θ_q is efficiency score of the DMU$_q$. The efficiency measure of the model (5) is always lower or equal to 1 and it is possible to simply prove that it is a product of efficiency measures of two single stages given by model (3) with constant returns to scales - see e.g. [6] or [1]. Target values for inputs, intermediate characteristics and final outputs of the inefficient DMUs, i.e. characteristics of virtual units, can be given as linear (convex) combination of DMUs using their optimal weights λ^*_j and μ^*_j, $j = 1,2,\dots,n$.

Another formulation of two-stage DEA model under constant returns to scale assumption is given in [4]. Their formulation follows:

$$
\begin{aligned}
&\text{Minimize} && \theta_q - \phi_q && \\
&\text{subject to} && \sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_q x_{iq}, && i = 1,2,\dots,m, \\
& && \sum_{j=1}^{n} y_{kj}\lambda_j \ge \tilde{y}_{kq}, && k = 1,2,\dots,r, \\
& && \sum_{j=1}^{n} y_{kj}\mu_j \le \tilde{y}_{kq}, && k = 1,2,\dots,r, \\
& && \sum_{j=1}^{n} z_{lj}\mu_j \ge \phi_q z_{lq}, && l = 1,2,\dots,p, \\
& && \theta_q \le 1, \phi_q \ge 1, && \\
& && \lambda_j \ge 0, \mu_j \ge 0, && j = 1,2,\dots,n,
\end{aligned}
\tag{6}
$$

where λ_j and μ_j, $j = 1,2,\dots,n$, are weights of the DMUs in the first and second stage, θ_q and φ_q 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 (4) 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. Target values for inputs, intermediate characteristics and final outputs of the inefficient DMUs can be in the same way as in the previous model. The inefficient units in model (6) can be ranked relatively by the following geometric average efficiency measure:

$$e_q = (\theta_q / \varphi_q)^{1/2}. \tag{7}$$

Bank	Equity	FTE	Credits	Deposits	Profit
CS	14 014	10 163	416 854	530 101	14 317
Citibank	673	900	125 062	169 425	14 310
ČSOB	3 635	6 420	209 172	568 199	13 572
GE Money	775	2 290	97 262	109 942	3 851
Hypotecní Bank	157	476	145 070	455	2 288
ING CZ	84	123	19 169	89 211	821
KB	6 556	7 883	334 834	441 285	14 417
LBBW Bank	727	350	19 161	18 929	50
Raiffeisenbank	989	528	152 663	125 936	2 320
UniCredit	1 578	265	172 070	174 373	3 473
Volksbank CZ	186	726	39 147	30 155	345
Wustenrot	380	200	31 978	35 956	341
Commerzbank	42	371	40 162	19 950	343
Czech Export B	39	144	59 856	58 690	217
Czech-Moravian	171	218	20 776	27 076	1 053
PPF Bank	44	154	18 655	36 332	804
Blue pyramid	399	356	49 030	69 119	1 026
Raiffeisen SS	56	257	39 578	76 160	830
CS SS	420	250	44 307	97 540	1 464
Wustenrot hyp	17	33	11 261	2 259	90

Table 1 Data set – efficiency evaluation of Czech banks

The models (5) and (6) are models that measure efficiency of two-stage serial production processes under the assumption of constant returns to scale. They can be modified for variable returns to scale by adding convexity constraints for both sets of weights λ_j and μ_j, $j = 1, 2,\ldots, n$. The models (5) and (6) can be simply generalized for processes with more than two stages – see e.g. [1].

3 Efficiency evaluation of Czech banks

Applications of two- or multiple-stage DEA models are numerous. Results of the above formulated models are illustrated in this section on efficiency evaluation of 20 banks operating on the Czech financial market. It is rather an illustrative example than a serious case study even though the data set used in the example has real background. They are taken from public financial statements of the banks for year 2010. The first stage evaluates production efficiency and the second stage is profit efficiency. The following inputs, intermediate characteristics (outputs of the first stage and inputs of the second one) and final outputs are taken into account (source values of the characteristics for all 20 banks are presented in Table 1):

Inputs:
- Equity of the bank in millions of CZK (Czech crowns),
- Number of full time employees (FTE).

Intermediate characteristics (outputs of the first stage and inputs of the second one):
- Deposits in millions of CZK,
- Credits in millions of CZK.

Final output:
- Profit in millions of CZK.

Bank	1st stage VRS-I	2nd stage VRS-O	Overall Geomean	Model (5) VRS-I	Model (5) VRS-O	Model (6) VRS - e_q
(1)	(2)	(3)	(4)	(5)	(6)	(7)
CS	1.000	0.993	0.997 (5)	0.0836 (19)	0.9930 (4)	0.2890 (15)
Citibank	1.000	1.000	1.000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)
ČSOB	1.000	0.946	0.972 (6)	0.1709 (14)	0.9446 (5)	0.4190 (7)
GE Money	0.464	0.407	0.434 (14)	0.0485 (20)	0.2691 (7)	0.4834 (5)
Hypotecní Bank	1.000	1.000	1.000 (1)	0.1716 (13)	0.3286 (6)	0.4179 (8)
ING CZ	1.000	0.761	0.873 (7)	0.3832 (7)	0.1773 (9)	0.4210 (6)
KB	1.000	1.000	1.000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)
LBBW Bank	0.150	0.046	0.084 (20)	0.0943 (18)	0.0045 (20)	0.0667 (20)
Raiffeisenbank	0.968	0.207	0.447 (13)	0.1336 (16)	0.1688 (10)	0.4109 (10)
UniCredit	1.000	0.243	0.493 (10)	0.3127 (9)	0.2427 (8)	0.4849 (4)
Volksbank CZ	0.159	0.132	0.145 (19)	0.0975 (17)	0.0468 (17)	0.2163 (17)
Wustenrot	0.399	0.127	0.225 (17)	0.1835 (12)	0.0429 (18)	0.2071 (18)
Commerzbank	0.716	0.182	0.361 (15)	0.4314 (6)	0.0656 (16)	0.2561 (16)
Czech Export Bank	1.000	0.043	0.207 (18)	0.4503 (5)	0.0428 (19)	0.1684 (19)
Czech-Moravian	0.302	0.823	0.499 (9)	0.2355 (10)	0.1584 (12)	0.3980 (11)
PPF Bank	0.688	0.793	0.739 (8)	0.4581 (4)	0.1525 (13)	0.3905 (12)
Blue pyramid	0.373	0.213	0.282 (16)	0.1392 (15)	0.1133 (15)	0.3367 (14)
Raiffeisen SS	1.000	0.229	0.478 (11)	0.3620 (8)	0.1502 (14)	0.3503 (13)
CS SS	0.639	0.347	0.471 (12)	0.2227 (11)	0.1716 (11)	0.4143 (9)
Wustenrot hyp	1.000	1.000	1.000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)

Table 2 Efficiency scores given by DEA models

The results and comparison of presented models for all DMUs are included in Table 2. All the results are given by models under the assumption of variable returns to scale which seems to be more suitable for evaluation purposes. The table contains the following information:

- Column (1) – name of the bank.
- Column (2) – efficiency score of the first stage given by standard BCC input oriented model. As presented, half of the banks are determined as efficient by this model in the first stage.
- Column (3) – efficiency score of the second stage given by BCC output oriented model. As the efficiency scores given by output oriented models are greater or equal 1, their reciprocal values are presented because it is more understandable for decision makers. Only 4 banks are recognized as efficient by this model.
- Column (4) – geometric average of efficiency scores in the previous two columns and ranking of DMUs by this value (in parenthesis). Four units (Citibank, Hypotecni Bank, KB, and Wustenrot hyp) are efficient according to this measure.
- Column (5) – efficiency scores computed by two-stage model (5) with input orientation and variable returns to scale assumption. The results of this model are quite surprising. Some of the DMUs efficient or nearly efficient by two single stages models are rated very badly in this model. It holds especially for first DMU (CS). The reason is its low final output (profit) comparing to its inputs (equity and FTE) and inputs and profit of other units of the decision set.
- Column (6) – the same characteristics as in the previous column but the model (5) with output orientation is used. The efficiency scores are given as their reciprocal values as in column (3). The first DMU is almost efficient by this model and it is immediately after the first three efficient units with maximum level of efficiency. Differences in rankings of DMUs by input- and output oriented models are significant and they are hardly possible to explain.
- Column (7) – efficiency scores given by model (6) with variable returns to scale transformed using formula (7). The model (6) evaluates the efficiency in two stages simultaneously and input orientation in the first stage and output orientation in the second one is applied.

The results (efficiency score) of the model (5) are questionable. The efficiency score is a product of two efficiency scores from both stages with the same orientation and the same returns to scale assumption. This “overall” efficiency score can be simply used for ranking of DMUs but has no direct explanation. A special attention must be given to the orientation of the model because when applying the model the results are quite different depending on the orientation. The model (6) evaluates the two-stage efficiency simultaneously without necessity to choose a model orientation. It is its advantage but the interpretation of the results is problematic as in the previous case. A promising research direction consists in formulation of two- or multi-stage serial DEA models based on slack based DEA models. The model formulated in [9] by Tone seems to be an ideal starting point for experiments in this area.

4 Conclusions

Evaluation of efficiency of network production systems is a very complex task that can be solved using network DEA models. This class of models is widely discussed within professional community in the last years and many papers are published with a main focus on theoretical aspects and/or applications in this field. This paper is focused on a simplest system which is two-stage serial model. Two of main two-stage serial DEA models are formulated and applied on a real data set regarding banks operated in the Czech Republic. Efficiency of banks is evaluated by two-stage model where the first stage takes into account production efficiency and the second one profit efficiency. The conclusions given by applied models show quite significant differences in ranking of banks. It is difficult to explain the results (efficiency scores) and it is one of the most important disadvantages of current two-stage models. Further research will be focused on formulation of special network models with better stability of results and their clearer interpretation. Other interesting area is analysis of super-efficiency of DMUs in network models.

Acknowledgements

The research is supported by the Czech Science Foundation, project no. 402/12/1387, and by IGA FIS VSE Praha, project no. F4/18/2011.

References

[1] Ashrafi, A., Jaafar, A.B., Lee, L.S. and Abu Bakar, M.R., The efficiency measurement of parallel production systems: a non-radial DEA model. *Journal of Computer Science*, **7**, 5, 749-756 (2011).

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

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

[4] Chen, Y., Liang, L. and Zhu, J., Equivalence in two-stage DEA approaches. *European Journal of Operational Research*, **193**, 600-604 (2009).
[5] Jablonsky, J. A two-stage AHP/DEA model for evaluation of efficiency. In: T. Subrt (Ed.), *Mathematical Methods in Economics 2009*, University of Life Sciences Press, Praha, 159 - 164 (2009).
[6] Kao, C. and Hwang, S.N., 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 (2008).
[7] Liang, L., Li, Z. Q., Cook, W.D. and Zhu, J., Data envelopment analysis efficiency in two-stage networks with feedback. *IIE Transactions*, **43**, 5, 309-322 (2011).
[8] Paradi, J.C., Rouatt, C. and Zhu, H., Two-stage evaluation of bank branch efficiency using data envelopment analysis. *Omega*, **39**, 1, 99-109 (2011).
[9] Tone, K.: A slack-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research*, **143**, 32-41 (2002).
[10] Zhu, J. *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*. Boston: Kluwer Academic Publishers (2003).