Trendy v podnikání, roč. 6, číslo 4, s. 36-42, 2016.
Business Trends, vol. 6, no. 4, p. 36-42, 2016.

RANKING MODELS IN DATA ENVELOPMENT ANALYSIS

Josef Jablonský[1]

[1] prof. Ing. Josef Jablonský, CSc., University of Economics, Prague, Faculty of Informatics and Statistics, jablon@vse.cz

Abstract: Traditional data envelopment analysis models split decision making units into two basic groups, efficient and inefficient. They are based on solving linear optimization problems and currently they represent very popular tool for efficiency and performance evaluation. Efficiency scores of inefficient units allows their ranking but efficient units cannot be ranked directly because of their maximum efficiency score. The paper presents the most popular ranking models (Andersen and Petersen model, Tone's SBM model and cross efficiency evaluation) and discusses their advantages and drawbacks. All presented models are illustrated on an example with academic background - evaluation of academic departments that consumes several resources (academic staff and sum of salaries) and produces outputs as direct or indirect teaching, research outputs, etc. Computational aspects of all presented models are discussed.

Keywords: Data envelopment analysis, efficiency, super-efficiency, ranking

JEL Classification: C65

INTRODUCTION

Data envelopment analysis (DEA) has become a very popular tool for efficiency and performance analysis based on solving simple linear programs. DEA history starts by formulation of the first model by Charnes et al. (1978) and DEA theory is still subject of interest of many researchers. DEA models analyze relative technical efficiency of the set of n homogenous decision making units (DMUs) that use m inputs and produce r outputs. The efficiency score θq of the DMUq is defined as the weighted sum of outputs divided by the weighted sum of inputs as follows:

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

where uk, k = 1,…,r, is the positive weight of the k-th output, vi, i = 1,…,m, is the positive weight of the i-th input, and xij, i = 1,…,m, j = 1,…,n, and ykj, k = 1,…,r, j = 1,…,n, are non-negative values of the i-th input and the k-th output for the DMUj respectively. Traditional DEA models maximize the efficiency score (1) under the assumption that the efficiency scores of all other DMUs do not exceed 1 (100%). This problem must be solved for each DMU separately, i.e. in order to evaluate the efficiency of all DMUs the set of n optimization problems must be solved. The presented problem is not linear in objective function but it can be converted into a linear optimization problem and then solved easily. The transformation consists in maximization of the nominator or minimization of the denominator in expression (1). The constraints of this linear optimization problem express the upper bound for efficiency scores of all DMUs except the unit q and the unit sum of the denominator/nominator in (1). The model that maximizes the nominator in (1) is referenced as DEA input oriented model, the model that minimizes the denominator is DEA output oriented model. In both cases the DMUs with θq = 1 are lying on the efficient frontier estimated by the model and denoted as efficient units. Otherwise the units are inefficient and the efficiency score can be explained as a rate of input reduction or output expansion in order to reach the maximum efficiency. In some cases, it is more convenient working with dual problem to the linearized version of the model described above. This model, often referenced as input oriented envelopment CCR (Charnes, Cooper and Rhodes) model is formulated as follows:

minimize θ_q^{CCR}

subject to

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

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

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

where λ_j, $j = 1,..,n$, are the weights of the DMUs, and s_i^-, $i = 1,\ldots,m$, and s_k^+, $k = 1,\ldots,r$, are slack and surplus variables belonging to the inputs and outputs respectively. The optimal objective value of model (2) is lower than 1 for inefficient units and equals 1 for the units weakly or fully efficient. In order to rank efficient units many models based on various principles have been proposed. The inefficient units can be ranked easily according to their efficiency scores (lower score indicates lower ranking) but the efficient ones cannot be ranked directly due to their maximum identical efficiency score. The aim of this paper is to present main models for ranking of efficient units in DEA models and discuss their advantages and drawbacks. The results given by the ranking models will be illustrated on a small numerical example with an academic background.

1. RANKING OF EFFICIENT UNITS IN DEA

The efficiency score in CCR model (2) is limited to 1 and reaches its maximum values for efficient units. The number of efficient units cannot be estimated beforehand but one can easily imagine that this number can be relatively high and even, in special cases, all units in the given set may be identified by DEA models as efficient. This situation may occur especially in cases when the number of evaluated units is quite low comparing to the total number of variables (inputs and outputs). The conclusion that all or almost all units are efficient is not helpful for decision makers at all. That is why at least a tool for a discrimination among efficient units may be helpful. DEA theory offers many approaches based on various methodological principles for this purpose. Their overview can be found e.g. in (Jablonsky, 2012). The most important category of DEA ranking models is represented by super-efficiency models. This class of models removes the evaluated unit from the set of DMUs and measures its distance from the new efficient frontier. In super-efficiency models the efficiency scores of inefficient units remain unchanged but the efficiency scores of efficient units are higher than 1. The efficient units can be simply ranked according to their super-efficiency scores. Two important super-efficiency models – Andersen and Petersen model and Tone's super SBM model – are presented below.

Andersen and Petersen model (AP model)

In general, AP model is historically the first super-efficiency DEA model. It was formulated by Andersen and Petersen (1993). Its input oriented formulation (3) is very close to the traditional input oriented formulation of CCR model (2). In this model the weight of the DMUq q is set to zero. It causes that the DMUq is removed from the set of units and the efficient frontier changes its shape after this removal. Super-efficiency score measures the distance of the evaluated DMUq from the new efficient frontier. Its value expresses how many times the inputs may increase (it means how they can get worse) in order the evaluated unit remains efficient. The AP model is as follows:

$$\begin{aligned}
&\text{minimize} && \theta_q^{AP} && \\
&\text{subject to} && \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta_q^{AP} x_{iq}, && i = 1,\dots,m, \\
& && \sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{kq}, && k = 1,\dots,r, \\
& && \lambda_j \geq 0, && j = 1,\dots,n,\ j \neq q, \\
& && \lambda_q = 0.
\end{aligned} \tag{3}$$

AP model has many drawbacks. Probably the most significant one consists in a possible infeasibility of model (3) under the assumption of non-constant returns to scale (the model is extended by additional constraint that limits the sum of all variables). In addition, the results given by this model are often hardly explainable.

Tone's super SBM model (super SBMT model)

Tone (2001) has proposed a slacks based measure of efficiency (SBM model) that measures the efficiency of the units under evaluation using slack variables only. This model that is basis for a super-efficiency SBM model presented in Tone (2002). The super-efficiency SBM model (super SBMT model) removes the evaluated unit DMUq from the set of units and looks for a DMU* with inputs xi*, i = 1,...,m, and outputs yk*, k = 1,...,r, being SBM (and CCR) efficient after this removal. It is clear that all inputs of the unit DMU* have to be greater or equal than inputs of the unit DMUq and all outputs will be lower or equal comparing to outputs of DMUq. The super-efficiency measure is the distance of two units DMUq and DMU* in their input and output space. Super SBMT model is formulated as follows:

$$\text{minimize} \quad \theta_q^{SBM} = \frac{\frac{1}{m}\sum_{i=1}^{m} x_i^* / x_{iq}}{\frac{1}{r}\sum_{k=1}^{r} y_k^* / y_{kq}}, \tag{4}$$

$$\begin{aligned}
&\text{subject to} && \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = x_{iq}, && i = 1,\dots,m, \\
& && \sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{kq}, && k = 1,\dots,r, \\
& && x_i^* \geq x_{iq}, && i = 1,\dots,m, \\
& && y_k^* \leq y_{kq}, && k = 1,\dots,r, \\
& && \lambda_j \geq 0, && j = 1,\dots,n,\ j \neq q, \\
& && \lambda_q = 0.
\end{aligned} \tag{5}$$

The numerator in ratio (4) can be explained as the distance of units DMUq and DMU* in input space and the average reduction rate of inputs of DMU* to inputs of DMUq. The same holds for output space in the denominator of ratio (4). The model (4)-(5) takes into account inputs and outputs and measures the distance in input and output space simultaneously. The objective function of the model is not linear but the model can be simply re-formulated as standard LP problem using Charnes-Cooper transformation. The super SBMT model (4)-(5) returns optimal objective value greater or equal 1. The optimal efficiency score is greater than 1 for efficient DMUs – higher value is assigned to more efficient units. All SBM inefficient units reach optimal score 1 in super SBMT model. That is why this model cannot be used for classification of inefficient units. In the first stage it is necessary to use any of traditional DEA models in order to identify inefficient units and then the SBMT model can be applied. In this way the decision maker can receive complete ranking of all units.

SBMT model is currently one of the most popular models for ranking of efficient units. Its advantage consists is simultaneous considering of inputs and outputs in calculation of final super-efficiency scores. Moreover, the results can be explained very easily to decision makers.

Cross efficiency evaluation

Cross efficiency evaluation is completely different concept which allows ranking of DMUs in DEA models comparing to super-efficiency models presented above. The basic idea of this concept is evaluation of each DMU using optimal weights of inputs and outputs of all DMUs given by a traditional DEA model. Let us suppose that the DMUj is the unit under evaluation and denote ukj, k = 1,…, r, j = 1,…, n optimal weights of the k-th output, and vij, i = 1,…, m, j = 1,…,n optimal weights of the i-th input given by the CCR input oriented model (6) in its multiplier form which is dual model to model (2). It is formulated as follows:

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

subject to (6)

$$\sum_{i=1}^{m} v_{iq} x_{iq} = 1,$$
$$\sum_{k=1}^{r} u_{kq} y_{kj} - \sum_{i=1}^{m} v_{iq} x_{ij} \le 0, \; j = 1,\dots,n,$$
$$u_{kq}, v_{iq} \ge \varepsilon, \qquad k = 1,\dots,r, i = 1,\dots,m.$$

Cross efficiency of the unit DMUq by using optimal weights given by model (6) in evaluation of DMUj is defined as follows:

$$E_{qj} = \frac{\sum_{k}^{r} u_{kj} y_{kq}}{\sum_{i}^{m} v_{ij} x_{iq}}, \qquad q = 1,\dots,n, j = 1,\dots,n. \tag{7}$$

It is clear that E_{qq} = θ_q^{CCR}, i.e. by using optimal weights of the unit DMUq its cross efficiency is equal to the efficiency score given by models (2) or (6). It can be easily proved that Eqj ≤ 1 for CCR input oriented models. Average cross efficiency φ_q is defined as follows:

$$\varphi_q = \frac{\sum_{j=1}^{n} E_{qj}}{n}, \text{ or } \varphi_q = \frac{\sum_{j=1, j \ne q}^{n} E_{qj}}{n-1}, \; q = 1,\dots,n. \tag{8}$$

Table 1: Data set

Department	Staff	Budget	Direct	Indirect	Research
D1	7	2709	1478	704	156
D2	13	8706	4028	1676	666
D3	4	1877	651	274	298
D4	15	11839	7091	5030	289
D5	11	5567	3106	999	314
D6	13	5299	3810	1021	215
D7	12	6513	3899	2192	303
D8	16	10193	6743	1948	486
D9	4	1478	1930	748	5

Source: own processing

Average cross efficiencies are positive and lower or equal than 1 but it is clear that the upper bound can appear in very special cases only. That is why the values q can be used for complete ranking of all DMUs. More information about cross efficiency models can be found in Sexton et al. (1986) and Green et al. (1996). Main drawback of cross efficiency evaluation is in a possible not uniqueness of optimal solution of model (6). This fact may lead to quite contradictory final results. In order to overcome this drawback many modifications of the presented main idea have been proposed – see e.g. Wang and Chin (2010). Another important drawback consists in impossibility to explain q values to decision makers. It is just a number that allows ranking of the units.

2. NUMERICAL ILLUSTRATION

The presented models in the previous section will be illustrated on an example dealing with performance evaluation of academic departments of a faculty. Even it is rather an illustration the data set presented below is based on a real situation in the faculty where the author is affiliated. Let is suppose the set of 9 departments (D1 to D9), each of them described by two inputs (the number of academic staff and the annual budget of the department in thousands of CZK), and by three outputs (direct teaching hours, indirect teaching hours and the number of publication points according to the methodology of the faculty). The complete data set is presented in Table 1.

The results of selected DEA models for the given data set are presented in Table 2. First column of the table contains efficiency scores of all departments calculated using CCR model (assumption of constant returns to scale). Five of nine units are efficient, the remaining ones are inefficient – e.g. D1 is not efficient and its efficiency score is 0.765, i.e. this department needs increase outputs or reduce inputs by 23,5 % in order to reach the efficient frontier. The worse unit is D6 with the efficiency score 0.744.

In order to discriminate among five efficient units the following ranking models have been applied:

- Andersen and Petersen super-efficiency model (AP),
- Tone's SBM super-efficiency model (SBMT),
- Cross-efficiency model that calculates q values (8) using the entire set of DMUs including inefficient units (Cross1), and
- Cross-efficiency model that calculates q values (8) using the set of efficient DMUs only (Cross2).

Table 2: Results

Dept.	CCR	AP	Rank	SBMT	Rank	Cross1	Rank	Cross2	Rank
D1	0.765	x	8	x	8	0.566	8	x	8
D2	1.000	1.021	5	1.011	5	0.731	4	0.637	4
D3	1.000	2.075	1	1.765	1	0.830	1	0.694	2
D4	1.000	1.810	3	1.406	3	0.807	3	0.695	1
D5	0.792	x	7	x	7	0.598	7	x	7
D6	0.744	x	9	x	9	0.557	9	x	9
D7	0.893	x	6	x	6	0.704	5	x	6
D8	1.000	1.045	4	1.028	4	0.700	6	0.628	5
D9	1.000	1.915	2	1.569	2	0.817	2	0.671	3

Source: own processing

The results in Table 2 show that ranking generated by both super-efficiency models is identical. It is not so surprising because they are based on same or very similar principles even it is not a rule that both models returns same rankings. Super-efficiency scores generated by SBMT model are always lower than the same characteristics given by AP model. It is clear that the first three units in final ranking (D3, D4 and D9) are rated significantly better than the remaining two efficient units (D2 and D8). Cross1 column in Table 2 contains average cross-efficiency scores calculated over all nine units (efficient and inefficient), i.e. the values in Table 2 are averages of nine cross-efficiency scores (7). In the contrary, Cross2 column takes into account just CCR efficient units, i.e. the values are averages of five cross-efficiency scores in our case. Even in Cross2 model there is rank reversal in first three places comparing to other models the remaining ranking is almost identical. The final cross-efficiency ranking is very close to the ranking given by super-efficiency scores but it is the situation of our example only. In general, both rankings may differ very significantly.

CONCLUSIONS

DEA models are applied in various areas of economic life. Efficiency scores that are limited by 1 and target values of inputs and outputs in order to reach efficient frontier are the main results that traditional DEA models offer to decision makers as information about all units under evaluation. An important information that is often required by decision makers is complete ranking of all DMUs. Inefficient DMUs are easily ranked according to their efficiency scores but the efficient ones cannot be ranked in this way. That is why a class of DEA models that allows their ranking has been formulated in the past. The most important group among DEA ranking models is the group of super-efficiency models. In their nature they are based on similar principles and that is why they usually return rankings very close each other but they differ in explanation of generated super-efficiency scores and other results. Two super-efficiency models were presented but several other have been proposed by researchers, an original super-efficiency model based on goal programming methodology is introduced in Jablonsky (2012). The research in this field is currently focused on ranking of DMUs in network DEA models or multi-period DEA models.

ACKNOWLEDGEMENTS

The research is supported by the Czech Science Foundation, project no. P402/12/G097 – Dynamic Models in Economy

REFERENCES

Andersen, P. and Petersen, N.C. (1993) A procedure for ranking efficient units in data envelopment analysis. *Management Science*, 39(10), 1261-1264.

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.

Green, R.H., Doyle, J.R. and Cook, W.D. (1996) Preference voting and project ranking using DEA and cross-efficiency evaluation.

European Journal of Operational Research, 90(3), 461-472.

Jablonsky, J. (2012) Multicriteria approaches for ranking of efficient units in DEA models. *Central European Journal of Operations Research*, 20(3), 435-449.

Sexton, T,R,, Silkman, R.H. and Hogan A.J. (1986) Data envelopment analysis: Critique and extensions. In: Silkman, R.H. (ed.) *Measuring efficiency: An assessment of data envelopment analysis*. San Francisco: Jossey-Bass.

Tone, K. (2001) A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, 130 (3), 498-509.

Tone, K. (2002) A slacks-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research*, 143 (1), 32-41.

Wang, Y.M. and Chin, K.S. (2010) Some alternative models for DEA cross-efficiency evaluation. *International Journal of Production Economics*, 128(1), 332-338.