# COMPARISON BETWEEN DATA ENVELOPMENT ANALYSIS MODELS WITH PENALTIES

Petra Zýková[1]

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

## Abstract

The paper deals with Data Envelopment Analysis (DEA) models with advanced voting systems for ranking of candidates with penalties. The main aims of the system are to find a general winner and ranking of all candidates. Every voter gives the ranking of the first t-candidates and can give penalties to candidates who he/she surely does not want to vote for. Advanced voting systems are being used based on the use of data envelopment analysis models. The original contribution of the paper consists in the formulation of a new DEA/AR model with penalties. This model is derived from the DEA/AR model with penalties. The proposed models are illustrated on a simulated data set. This paper aims to compare DEA models with penalties.

Keywords: data envelopment analysis, voting systems, penalty

## INTRODUCTION

Often, decision-makers need to find a general winner based on the ranking of candidates in several voting rounds. Less often, they have information that any voters do not want to vote for any candidates. This problem can be solved in many different ways. One of the possibilities is to apply data envelopment analysis (DEA) models. DEA models have been first developed by Charnes *et al.* (1978) based on the concept introduced by Farrell (1957). Using DEA models, the optimal vector of weights for every candidate can be derived. Therefore settings of weights cannot be influenced by the decision-maker. This use of DEA models was shown in Hashimoto (1997). The usage of DEA models with penalties was firstly developed in Zýková (2017) and Dlouhý, Jablonský, Zýková (2018). There is an application of DEA models with penalties on Formula 1 championship results in Zýková (2017). I have worked out the equal results

by DEA models without penalties and also with penalties. I have supposed that the result was caused by the data set (not enough penalties). This was the reason why I have generated random data set; on whose I have tested these DEA models. Again I have determined that penalties do not have any influence on the final ranking. That is why I have decided to modify DEA models with penalties. I wanted a new model with a bigger influence of penalties. And this new model is developed in this paper Models mentioned in this paper are illustrated on a simulated data set.

The paper is organized as follows. The next section presents a general familiarization with DEA models, the formulation of the problem and is paying attention to the DEA/AR models with penalties for the ranking of candidates. The following section consists of a numerical illustration of the DEA/AR models proposed in this paper. The final section of the paper summarises the results.

## MATERIALS AND METHODS

### DEA Models

DEA models are a general tool for efficiency and performance evaluation of the set of homogenous DMUs that spend multiple $w$ inputs and transform them into multiple $t$ outputs. The measure of efficiency (efficiency score) of this transformation is one of the main results of an application of DEA models. Let us denote $Y = (y_{rj}, r = 1, ..., t, j = 1, ..., n)$ a positive matrix of outputs and $X = (x_{kj}, k = 1, ..., w, j = 1, ..., n)$ a positive matrix of inputs. The efficiency score of the unit under evaluation $DMU_{j_0}$ is as follows:

Maximize
$$U_{j_0} = \frac{\sum_{r=1}^{t} u_r y_{rj_0}}{\sum_{k=1}^{w} v_k x_{kj_0}}$$
(1)

subject to
$$\frac{\sum_{r=1}^{t} u_r y_{rj}}{\sum_{k=1}^{w} v_k x_{kj}} \leq 1, \quad j = 1,...,n,$$

$$u_r \geq \varepsilon, \qquad r = 1, ..., t,$$

$$v_k \geq \varepsilon, \qquad k = 1, ..., w,$$

where $u_r$ is a positive weight of the $r$th output, $v_k$ is a positive weight of the $k$th input, and $\varepsilon$ is an infinitesimal constant. Model (1) is not linear in its objective function but may easily be transformed into a linear program. The linearized version of the input-oriented model (often called the CCR model) is as follows:

Maximize
$$U_{j_0} = \sum_{r=1}^{t} u_r y_{rj_0}$$

$$\sum_{k=1}^{w} v_k x_{kj_0} = 1,$$
(2)

subject to
$$\sum_{r=1}^{t} u_r y_{rj} - \sum_{k=1}^{w} v_k x_{kj} \leq 0, j = 1,...,n,$$

$$u_r \geq \varepsilon, \qquad r = 1, ..., t,$$

$$v_k \geq \varepsilon, \qquad k = 1, ..., w.$$

### Formulation of the Problem

The problem we are dealing with can be defined as follows: Let us assume a voting system with a ranking of $n$ candidates and $m$ voters. Every voter gives a ranking of the first t-candidates. This means that he/she has not to assign a ranking to all candidates but only to the top t-candidates ($t \leq n$).

A voter must not give the same ranking to any two candidates. The aim is to find a general winner and general ranking of all candidates. The general ranking of candidates is obtained according to points, which $j$-th candidate gets from voters:

$$s_j = \sum_{r=1}^{t} u_r y_{rj}, j = 1,...,n,$$

where $u_r$, $r = 1, ..., t$ is the number of points gained for the $r$th place, $y_{rj}$, $r = 1, ..., t$ and $j = 1, ..., n$, says how many times the $j$th candidate was in the $r$th place. We assume that the values $y_{rj}$ are known. The general winner is the candidate with the highest value of $s_j$.

In addition to the original formulation of the problem above, every voter can define negative candidates that are not acceptable for him/her. The aim is to find a general winner and a general ranking of all candidates. Let us assign the index $(t + 1)$ and the weight $u_{t+1}$ to the negative ranking (penalty).

### DEA Models for Ranking of Candidates

DEA models for ranking of candidates can be used in several ways. In the beginning, was DEA models without penalties. One of these models is shown below:

Maximize
$$h_{j_0} = \sum_{r=1}^{t} u_r y_{rj_0}$$

$$\sum_{r=1}^{t} u_r y_{jr} \leq 1, j = 1,...,n,$$
(3)

subject to
$$u_r - u_{r+1} \geq \varepsilon, \quad r = 1, ..., t - 1,$$

$$u_r - 2u_{r+1} + u_{r+2} \geq 0, r = 1, ..., t - 2,$$

$$u_t = \varepsilon.$$

Model (3) was derived in Zýková (2019) from DEA/AR models by Hashimoto (1997). More about DEA/AR models without penalties is in Zýková (2018), and also in Dlouhý, Jablonský, Zýková (2018).

First DEA/AR model with penalties was derived in Zýková (2017):

Maximize
$$h_{j_0}^- = \sum_{r=1}^{t+1} u_r y_{rj_0}$$

$$\sum_{r=1}^{t+1} u_r y_{jr} \leq 1, \quad j = 1,...,n,$$

$$u_r - u_{r+1} \geq \varepsilon, \quad r = 1, ..., t - 1,$$
(4)

subject to
$$u_t \geq \varepsilon,$$

$$\sum_{r=1}^{t+1} u_r y_{yj} \geq -1, \quad j = 1,...,n,$$

$$u_{t+1} \leq -\varepsilon.$$

We proposed a new model (5), where penalties should have more impact. After some testing, we have decided, that the weight of penalties may be equal or less to the average of the highest and the lowest weight. The new model is as follow:

Maximize
$$g_{j_0}^- = \sum_{r=1}^{t+1} u_r y_{rj_0}$$

$$\sum_{r=1}^{t} u_r y_{jr} \leq 1, \quad j = 1,...,n,$$

$$u_r - u_{r+1} \geq \varepsilon, \quad r = 1, ..., t - 1,$$

(5)

subject to
$$u_t \geq \varepsilon,$$

$$u_r - 2u_{r+1} + u_{r+2} \geq 0, r = 1, ..., t - 2,$$

$$u_{t+1} \leq -\frac{u_1 + u_t}{2}.$$

The model (5) differs from the model (4) in several ways. Firstly, the sum

$$\sum_{r=1}^{t} u_r y_{jr} \leq 1$$

is only from 1 to $t$. Secondly, we assume increasing and also the convex vector of weights. Thirdly, we have a new condition that defines the weight of penalties, which was mentioned before.

### RESULTS

We have simulated data for this paper by the uniform distribution. We assume a voting system with the ranking of 11 candidates and 21 voters. Epsilon is equal to $10^{-4}$. Results of voting are presented in Tab. I. The column "-" shows the number of penalties and column sum the total number of the points-scoring position for every candidate.

We have solved model (3), (4) and (5) using the originally written procedure in Lingo solving system.

Results of the models (3), (4) and (5) are presented in Tab. II. We are interested in the fight between 1st and 2nd candidate. Both have the same number of points-scoring positions. 1st candidate has more victories, but also more penalties. According to the model (3), both models are in the first place. We could solve this trouble, for example, using by DEA/AR exclusion model by Hashimoto (1997). Model (4) determines the 2nd candidate as the winner. The less number of penalties causes it. The difference is only one penalty, we see in Tab. II, the difference is only 0.0002. Nevertheless, the difference by using model (5) is much higher. It is caused by the bigger weight for the penalty.

### DISCUSSION

We can see the difference between the model (4) and (5) in the placing and values of the 6th candidate. 6th candidate has four penalties, which leave him to the worse placement. The weights of this candidate are presented in Tab. III and in Fig. 1. We can see the more significant influence of penalties in the model (5) compare to model (4). The influence of the penalties in the model (4) is low. We can see it in Tab. II, the ranking of the 6th candidate, is in both models 6th place. We could also study the difference between weights for other units.

I: *Data set – placement of candidates*

| Candidates | | | | | | Placement | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | - | sum |
| 1 | 5 | 2 | 1 | 4 | 2 | 2 | 0 | 0 | 3 | 4 | 4 | 23 |
| 2 | 3 | 3 | 2 | 4 | 3 | 2 | 3 | 0 | 0 | 3 | 3 | 23 |
| 3 | 3 | 1 | 2 | 1 | 2 | 1 | 6 | 1 | 3 | 1 | 3 | 21 |
| 4 | 3 | 1 | 1 | 2 | 1 | 4 | 0 | 0 | 4 | 3 | 3 | 19 |
| 5 | 3 | 0 | 0 | 4 | 4 | 1 | 1 | 3 | 1 | 0 | 2 | 17 |
| 6 | 1 | 4 | 2 | 0 | 1 | 2 | 2 | 5 | 1 | 0 | 4 | 18 |
| 7 | 1 | 3 | 3 | 0 | 1 | 0 | 4 | 3 | 1 | 0 | 2 | 16 |
| 8 | 1 | 3 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 0 | 2 | 15 |
| 9 | 1 | 0 | 5 | 2 | 3 | 2 | 0 | 5 | 3 | 3 | 4 | 24 |
| 10 | 0 | 3 | 2 | 0 | 0 | 2 | 2 | 2 | 3 | 2 | 4 | 16 |
| 11 | 0 | 1 | 2 | 2 | 2 | 4 | 1 | 0 | 1 | 5 | 3 | 18 |

Source: own processing

II: *Results of used DEA models – values of the objective function, placement*

| Candidates | model (3) | | model (4) | | model (5) | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| 1 | 1.0000 | 1 - 2 | 0.9992 | 2 | 0.8245 | 3 |
| 2 | 1.0000 | 1 -2 | 0.9994 | 1 | 0.8884 | 1 |
| 3 | 0.7823 | 3 | 0.8961 | 4 | 0.7630 | 4 |
| 4 | 0.6845 | 7 | 0.8131 | 5 | 0.6774 | 5 |
| 5 | 0.7293 | 5 | 0.7373 | 8 | 0.6494 | 6 |
| 6 | 0.7196 | 6 | 0.7743 | 6 | 0.6003 | 9 |
| 7 | 0.6614 | 8 | 0.6911 | 9 | 0.6028 | 8 |
| 8 | 0.6364 | 9 | 0.6493 | 11 | 0.5647 | 10 |
| 9 | 0.7610 | 4 | 0.9992 | 3 | 0.8331 | 2 |
| 10 | 0.4881 | 11 | 0.6742 | 10 | 0.4995 | 11 |
| 11 | 0.5295 | 10 | 0.7489 | 7 | 0.6243 | 7 |

Source: own processing

III: *Vectors of weights for the 6th candidate according to model (3), model (4) and model (5)*

|  | model (3) | model (4) | model (5) |
|:---:|:---:|:---:|:---:|
| u(1) | 0.10207 | 0.05051 | 0.04822 |
| u(2) | 0.08008 | 0.04787 | 0.04695 |
| u(3) | 0.05808 | 0.04523 | 0.04569 |
| u(4) | 0.03609 | 0.04259 | 0.04442 |
| u(5) | 0.03009 | 0.04186 | 0.04315 |
| u(6) | 0.02409 | 0.04113 | 0.04188 |
| u(7) | 0.01809 | 0.04041 | 0.04061 |
| u(8) | 0.01210 | 0.03968 | 0.03934 |
| u(9) | 0.00610 | 0.03895 | 0.03807 |
| u(10) | 0.00010 | 0.03822 | 0.03680 |
| u(11) |  | -0.00010 | -0.04251 |

Source: own processing
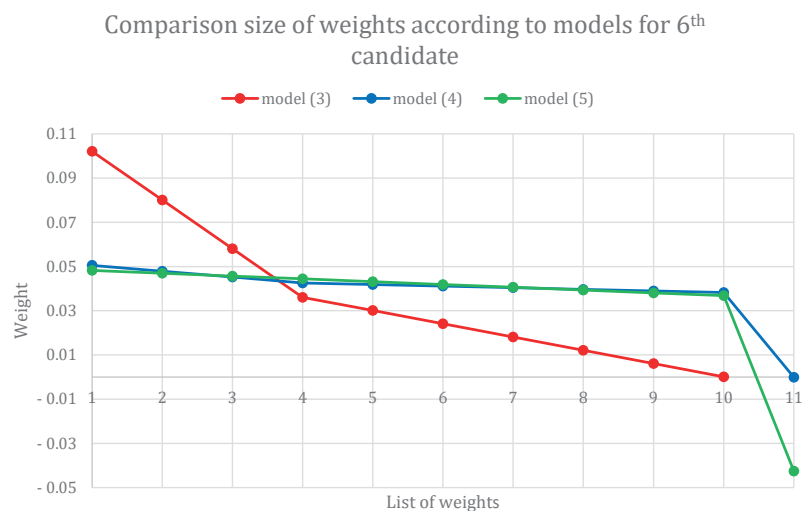


1: *Graph of vectors of weights for the 6th candidate according to model (3), model (4) and model (5)*
Source: own processing

## CONCLUSION

The determination of a general winner and complete ranking of candidates is an exciting task whose solution is significantly influenced by the vector of weights assigned to the candidates. This paper deals with the use of data envelopment analysis models with penalties that return a unique optimal vector of weights for particular candidates. As penalties, we think negative ranking, which every candidate can receive from voters. It means that voter this candidate does not want. This paper proposed a new DEA/AR models with penalties, in which penalties have a bigger influence than before proposed models. It was found out by calculation on a randomly generated data set. This model is the original contribution of the author of this paper. We compare this model with the older DEA/AR model with penalties and one model without penalties.

## REFERENCES

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.

DLOUHÝ, M., JABLONSKÝ, J. and ZÝKOVÁ, P. 2018. *Analýza obalu dat*. Praha: Professional Publishing.

FARRELL, M. 1957. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3): 253–290.

HASHIMOTO, A. 1997 A ranked voting system using a DEA/AR exclusion model: A note. *European Journal of Operational Research*, 97(3): 600–604.

ZÝKOVÁ, P. 2017. *Scoring rules – advanced voting systems with ranking of candidates*. Diploma Thesis. University of Economics, Prague.

ZÝKOVÁ, P. 2018. Data envelopment analysis models with penalties. In: VÁCHOVÁ, L. and KRATOCHVÍL, V. (Ed.). *Proceedings of the 36th International Conference on Mathematical Methods in Economics (MME 2018)*. Praha: MatfyzPress, Publishing House of the Faculty of Mathematics and Physics, Charles University, pp. 642–647.

ZÝKOVÁ, P. 2019. Properties of DEA models with assurance region. In: *The 8th International Conference on Modeling, Simulation and Applied Optimization*. Apr 15, 2019–Apr 17, 2019, University of Bahrain.

Contact information

Petra Zýková: petra.zykova@vse.cz