Josef Jablonsky

A SLACK BASED MODEL FOR MEASURING SUPER-EFFICIENCY IN DATA ENVELOPMENT ANALYSIS*

INTRODUCTION

Data envelopment analysis (DEA) is a tool for measuring the relative efficiency and comparison of decision making units (DMU). The DMUs are usually described by several inputs that are spent for production of several outputs. Let us consider the set E of n decision making units $E = \{DMU_1, DMU_2,\ldots, DMU_n\}$. Each of the units produces r outputs and spends m inputs for their production. Let us denote $\mathbf{x}^{\mathbf{j}} = \{x_{ij}, i = 1,2,\ldots,m\}$ the vector of inputs and $\mathbf{y}^{\mathbf{j}} = \{y_{ij}, i = 1,2,\ldots,r\}$ the vector of outputs of the DMU_j. Then $\mathbf{X}$ is the (m, n) matrix of inputs and $\mathbf{Y}$ the (r, n) matrix of outputs.

The basic principle of the DEA in evaluation of efficiency of the DMU_q, $q \in \{1,2,...,n\}$ consists in looking for a virtual unit with inputs and outputs defined as the weighted sum of inputs and outputs of the other units in the decision set – $\mathbf{X\lambda}$ a $\mathbf{Y\lambda}$, where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2,\ldots, \lambda_n)$, $\boldsymbol{\lambda} > 0$ is the vector of weights of the DMUs. The virtual unit should be better (or at least not worse) than the analysed unit DMU_q. The problem of looking for a virtual unit can generally be formulated as a standard linear programming problem:

$$\begin{aligned} &\text{minimise} && \theta \\ &\text{subject to} && \mathbf{Y\lambda} \geq \mathbf{y^q} \qquad (1)\\ & && \mathbf{X\lambda} \leq \theta\mathbf{x^q} \\ & && \boldsymbol{\lambda} \geq 0 \end{aligned}$$

The DMU_q is to be considered as efficient if the virtual unit is identical with evaluated unit (virtual unit with better inputs and outputs does not exist). In this case $\mathbf{Y\lambda} = \mathbf{y^q}$, $\mathbf{X\lambda} = \mathbf{x^q}$ and minimum value of $z = \theta = 1$. Otherwise

* The research is supported by the Grant Agency of Czech Republic – grant No 402/03/1360.

the DMU_q is not efficient and minimum value of $\theta < 1$ can be interpreted as the need of a proportional reduction of inputs in order to reach the efficient frontier. The presented model is input oriented model because its objective is to find a reduction rate of inputs in order to reach the efficiency. Analogously can be formulated output oriented model.

Model (1) shows just the basic philosophy of DEA models. The first DEA model was formulated in 1978 by Charnes, Cooper, Rhodes (CCR model). Its input oriented form (CCR-I) looks as follows:

$$\begin{aligned} &\text{minimise} && z = \theta - \varepsilon(\mathbf{e}^{\mathbf{T}}\mathbf{s}^{+} + \mathbf{e}^{\mathbf{T}}\mathbf{s}^{-}) \\ &\text{subject to} && \mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^{+} = \mathbf{y}^{\mathbf{q}} \\ & && \mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^{-} = \theta\mathbf{x}^{\mathbf{q}} \\ & && \boldsymbol{\lambda}, \mathbf{s}^{+}, \mathbf{s}^{-} \geq 0 \end{aligned} \tag{2}$$

where $\mathbf{e}^{\mathbf{T}} = (1,1,\ldots,1)$ and ε is a infinitesimal constant (usually 10^{-8}). Presented formulations (1) and (2) are very close each other. The variables $\mathbf{s}^{+}, \mathbf{s}^{-}$ are just slack variables expressing the difference between virtual inputs/outputs and appropriate inputs/outputs of the DMU_q. Obviously, the virtual inputs/outputs can be computed using the optimal values of variables of the model (2) as follows:

$$\mathbf{x}^{\mathbf{q}'} = \mathbf{x}^{\mathbf{q}}\theta^{*} - \mathbf{s}^{-}$$
$$\mathbf{y}^{\mathbf{q}'} = \mathbf{y}^{\mathbf{q}} + \mathbf{s}^{+}$$

The CCR model supposes constant returns to scale – it is supposed that a considered percentual change of inputs leads to the same percentual change of outputs. The modification of the CCR model taking into account variable returns to scale (so called BCC model) is derived from model (2) by adding the convexity constraint $\mathbf{e}^{\mathbf{T}}\boldsymbol{\lambda} = 1$.

1. SUPER-EFFICIENCY MODELS

The efficiency score in standard DEA models is limited to unity (100%). Nevertheless, the number of efficient units identified by DEA models and reaching the maximum efficiency score 100% can be relatively high and especially in problems with a small number of decision making units the efficient set can contain almost all the units. In such cases it is very important to have a tool for a diversification and classification of efficient units. That is why several DEA models for classification of efficient units were formulated. In these models the efficient scores of inefficient units remain lower than 100% but the efficiency score for efficient units can be higher than 100%. Thus the efficiency score can be taken as a basis for a complete ranking of efficient units. The DEA models that relax the condition for unit efficiency are called super-efficiency models.

Basic principles of super-efficiency models are illustrated on Figure 1. It presents an example with 8 DMUs, each of them described by one input and one output. The Figure shows the BCC efficient frontier defined by units DMU_3, DMU_2, DMU_8 and DMU_6. These units are BCC efficient and their efficiency score is equal to 1 because it is not possible to find any convex combination of other units with better characteristics (lower input and higher output). The remaining four units are not efficient. The super-efficiency models are always based on removing the evaluated efficient unit from the set of units (unit DMU_2 on Figure 1). This removal leads to the modification of the efficient frontier (heavy line of Figure 1) and the super-efficiency is measured as a distance between evaluated unit (DMU_2) and a unit on the new efficient frontier (DMU*). Of course several distance measures can be used – this leads to different super-efficiency definitions.

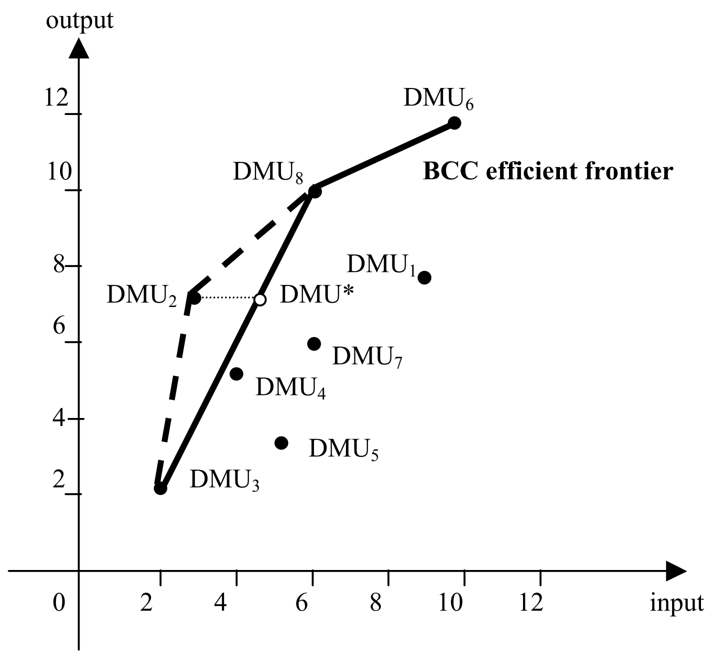

Fig. 1. A super-efficiency measure

The first super-efficiency DEA model was formulated by Andersen and Petersen [1]. Its input oriented formulation (3) is very close to the standard input oriented formulation of the CCR-I model (2). In this model the weight λ_q of the evaluated unit DMU_q is equated to zero. This cannot influence the efficiency score of the inefficient units but the efficiency score of the efficient units is not limited by unity in this case. The input oriented formulation of the Andersen and Petersen model is as follows:

$$\begin{aligned} &\text{minimise} && \theta \\ &\text{subject to} && \sum_{j=1,\neq q}^{n} x_{ij}\lambda_j + s_i^- = \theta x_{iq}, \qquad i = 1,2,...,m \\ & && \sum_{j=1,\neq q}^{n} y_{ij}\lambda_j - s_i^+ = y_{iq}, \qquad i = 1,2,...,r \\ & && \boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^- \geq 0 \end{aligned} \tag{3}$$

Tone [6] proposes a slack based measure of efficiency (SBM model) that is basis for his formulation of the super-efficiency model presented in [7]. The Tone's SBM model is formulated as follows:

$$\begin{aligned} &\text{minimise} && \rho = \frac{1 - \frac{1}{m}\sum_{i=1}^{m} s_i^-/x_{iq}}{1 + \frac{1}{r}\sum_{i=1}^{r} s_i^+/y_{iq}} \\ &\text{subject to} && \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = x_{iq}, \qquad i = 1,2,...,m \\ & && \sum_{j=1}^{n} y_{ij}\lambda_j - s_i^+ = y_{iq} \qquad i = 1,2,...,r \\ & && \boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^- \geq 0 \end{aligned} \tag{4}$$

The formulation shows that the SBM model is non-radial and deals directly with slack variables. The model returns efficiency score between 0 and 1 and is equal to 1 if and only if the DMU_q is on the efficient frontier without any slacks. It is possible to prove that the efficiency score of the SBM model is always lower or equal than the efficiency score of the appropriate CCR input oriented model. The formulation of the model (4) with fractional objective function can be simply transformed into a standard problem with linear objective function.

The super-efficiency SBM model removes the evaluated unit DMU_q from the set of units (like Andersen and Petersen model) and looks for a DMU* with inputs x* and outputs y* being SBM (and CCR) efficient after this removal. It is clear that all the inputs of the unit DMU* have to be higher or equal than the inputs of the unit DMU_q and all the outputs will be lower or equal comparing to outputs of DMU_q. The super-efficiency is measured as a distance of the inputs/outputs of both the units. As a distance measure in the mathematical formulation of the super SBM model below, the variable δ is used:

minimise
$$\delta = \frac{\frac{1}{m}\sum_{i=1}^{m} x_i^* / x_{iq}}{\frac{1}{r}\sum_{i=1}^{r} y_i^* / y_{iq}} \qquad (5)$$

subject to
$$\sum_{j=1,\neq q}^{n} x_{ij}\lambda_j + s_i^- = x_{iq} \qquad i = 1,2,...,m \qquad (6)$$

$$\sum_{j=1,\neq q}^{n} y_{ij}\lambda_j - s_i^+ = y_{iq}, \qquad i = 1,2,...,r$$

$$x_i^* \geq x_{iq}, \qquad i = 1,2,...,m$$

$$y_i^* \leq y_{iq}, \qquad i = 1,2,...,r$$

$$\boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^-, \mathbf{y}^* \geq 0$$

The numerator in the ratio (5) can be interpreted as a distance of both the units in the input space and an average reduction rate of inputs of DMU* to inputs of DMU_q. The same holds for the output space in the denominator of the ratio (5). The model (5)-(6) takes into account both the inputs and outputs and measures the distance in the input and output space simultaneously. It is not a model with linear objective function but it can be simply re-formulated as a standard LP problem by means of Charnes-Cooper transformation.

Similarly to the previous model the input (output) oriented modification can be formulated. This modified model measures the distance of the DMU_q and the DMU* in the input (output) space only. The formulation of the input (output) oriented SBM model is derived from the model (5)-(6) by setting the denominator equal to 1, i.e. $y_i^* = y_{iq}$ (setting the numerator equal to 1, i.e. $x_i^* = x_{iq}$). The input oriented formulation of the super SBM model (SBM-I) is given as follows (similarly can be written output oriented form SBM-O):

minimise $$\delta_I = \frac{1}{m}\sum_{i=1}^{m} x_i^*/x_{iq} \tag{7}$$

subject to
$$\sum_{j=1,\neq q}^{n} x_{ij}\lambda_j + s_i^- = x_{iq}, \quad i = 1,2,...,m$$
$$\sum_{j=1,\neq q}^{n} y_{ij}\lambda_j - s_i^+ = y_{iq}, \quad i = 1,2,...,r \tag{8}$$
$$x_i^* \geq x_{iq}, \quad i = 1,2,...,m$$
$$y_i^* = y_{iq}, \quad i = 1,2,...,r$$
$$\boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^- \geq 0$$

In the previous models it holds $\delta^*(DMU_q) \geq 1$ ($\delta_I^*(DMU_q) \geq 1$) where δ^* (δ_I^*) are the optimal objective function values of models (5)-(6) and (7)-(8). The optimal efficiency score is greater than 1 for SBM efficient DMUs and it is possible to prove that $\delta_I^*(DMU_q) \geq \delta^*(DMU_q) > 1$ – higher value is assigned to more efficient units. All the SBM inefficient units reach in the super SBM model optimal score 1. It means that this model cannot be used for classification of inefficient units and have to be used in two steps:

- to apply the SBM model (4) in order to identify efficient units and classify inefficient units,
- to compute the super-efficiency scores by means of one of the super--efficiency SBM models – the models (5)-(6) or (7)-(8).

2. THE SBMG MODEL

The super-efficiency SBM models measure the distance of the evaluated unit DMU_q from a virtual unit DMU* under the assumption that its inputs **x*** and outputs **y*** are not better than the inputs $\mathbf{x}^q$ and outputs $\mathbf{y}^q$ of the unit DMU_q. When this assumption is not kept the super-efficiency can be measured by the following goal programming model (super SBMG model):

minimise $\rho = 1 + t\gamma + (1-t)\left(\sum_{i=1}^{m}\left[s_{1i}^{+}/x_{iq}\right] + \sum_{i=1}^{r}\left[s_{2i}^{-}/y_{iq}\right]\right)$

subject to

$$\sum_{j=1,\neq q}^{n} x_{ij}\lambda_j + s_{1i}^{-} - s_{1i}^{+} = x_{iq}, \quad i = 1, 2, ..., m \tag{9}$$

$$\sum_{j=1,\neq q}^{n} y_{ij}\lambda_j + s_{2i}^{-} - s_{2i}^{+} = y_{iq}, \quad i = 1, 2, ..., r$$

$$s_{1i}^{+}/x_{iq} \le \gamma, \quad i = 1, 2, ..., m$$

$$s_{2i}^{-}/y_{iq} \le \gamma, \quad i = 1, 2, ..., r$$

$$t \in \langle 0,1 \rangle$$

$$\lambda_j \ge 0,\ s_{1i}^{-} \ge 0,\ s_{1i}^{+} \ge 0,\ s_{2i}^{-} \ge 0,\ s_{2i}^{+} \ge 0$$

In this model the distance between the units DMU_q a DMU* is measured by the positive and negative deviational variables – s_{1i}^{-}, s_{1i}^{+} for the inputs and s_{2i}^{-}, s_{2i}^{+} for the outputs. The minimised objective function ρ contains just the positive deviations for the inputs and the negative ones for the outputs, because the model tries to measure the distance in the undesirable way for both the groups of characteristics – for the inputs the undesirable deviations are the positive ones, for the outputs the negative ones on the contrary.

The goal programming technique can optimise the problem by using deviational variables in two basic ways – minimisation of the weighted sum of deviational variables and minimisation of the maximum deviation. That is why the objective function of the model (9) consists of two parts – the first one minimises the maximum relative deviation γ and the second one the sum of relative deviations from the inputs and outputs characteristics of the unit DMU_q. Depending on the selection of the parameter t the model either minimises the maximum deviation γ ($t = 1$) or the weighted sum of deviations ($t = 0$). By selection of parameter t between zero and one, $t \in (0,1)$, the both approaches can be combined.

The super-efficiency models have to fulfil the basic requirement – when any input or output of the evaluated unit DMU_q worsens (improves), its super-efficiency score have to decrease (increase) or at least remain without changes. The proof of this feature for Tone's models is given in [7]. The proof for the model (9) follows directly from its definition – the worsening of the i-th input (output) does not lead to the higher positive deviational variable s_{1i}^{+} (negative variable s_{1i}^{-}) and by this the super-efficiency score cannot be higher.

Let us denote ρ_0*(DMU$_q$) the super-efficiency score of the unit DMU$_q$ given by the model (9) with parameter t = 0 and ρ_1*(DMU$_q$) the score given by the model (9) with parameter t = 1. It is obvious that the score ρ_0*(DMU$_q$) is always greater or equal to the score ρ_1*(DMU$_q$). Both the characteristics are always lower than the super-efficiency score given by Andersen and Petersen model (3). The objective function of the super SBM model (5)-(6) is defined as the ratio of the average deviations. The objective function of the SBMG model (9) pro t = 0 is the sum of positive deviations of inputs and negative deviations of outputs. That is why it is possible simply to show that the following relation holds: δ*(DMU$_q$) ≥ 1 + (ρ_0*(DMU$_q$) – 1)/(m + r), where δ*(DMU$_q$) is the super-efficiency score given by the model (5)-(6).

The objective function of the model (9) for parameter t = 0 is the sum of the relative undesirable deviations. This sum can be replaced by the average of all the undesirable slacks. The results in this case can often be better explained. The model (9) makes it possible to perform a sensitivity analysis of the problem – according to the optimum dual values of the input and output constraints it is possible to find out how the changes of the input and output values of the evaluated unit influence the SBGM super-efficiency score.

3. A NUMERICAL ILLUSTRATION

The presented three concepts of super-efficiency, including our own definition (9), are illustrated on the small numerical example taken from [7]. The example considers 6 decision making units (power plants locations) with four inputs and two outputs defined as follows:

- manpower required ($\mathbf{x}^1$),
- estimated construction costs in millions of USD ($\mathbf{x}^2$),
- annual maintenance costs in millions of USD ($\mathbf{x}^3$),
- the number of villages that have to be evacuated ($\mathbf{x}^4$),,
- plant power in megawatts ($\mathbf{y}^1$),
- safety level given by an ordinal scale from 1 to 10 (higher values are better) - ($\mathbf{y}^2$).

The input and output data of the problem are given in Table 1.

Table 1

Data of the problem

	Inputs				Outputs	
	x^1	x^2	x^3	x^4	y^1	y^2
DMU_1	80	600	54	8	90	5
DMU_2	65	200	97	1	58	1
DMU_3	83	400	72	4	60	7
DMU_4	40	1000	75	7	80	10
DMU_5	52	600	20	3	72	8
DMU_6	94	700	36	5	96	6

Applying the DEA models to small set of DMUs with relation to the number of inputs and outputs of the problem can often lead to the result that all the units are efficient. This situation occurs in our illustrative example. When the decision maker wants to discriminate among the efficient units the super-efficiency DEA models can be used. The super-efficiency score computed by Andersen and Petersen model (3), Tone's SBM model (5)-(6) SBMG model (9) for t = 0 and t = 1 are presented in Table 2.

Table 2

Comparison of super-efficiency scores

	A-P θ^*	SBM δ^*	SBMG t=0, ρ_0^*	SBMG t=1, ρ_1^*
DMU_1	1.0283	1.0116	1.0275	1.0139
DMU_2	2.4167	1.4146	1.5862	1.4146
DMU_3	1.3125	1.0781	1.2976	1.1351
DMU_4	1.6250	1.1563	1.5556	1.2381
DMU_5	2.4026	1.5859	1.8454	1.4122
DMU_6	1.0628	1.0198	1.0591	1.0304

Table 2 illustrates a conclusion derived by Tone [7] that the super-efficiency score for unit DMU_q given by the A-P model (3) is always greater or equal to the score given by the SBM model, ie. $\theta^*(DMU_q) \geq \delta^*(DMU_q)$. Of course it holds $\theta^*(DMU_q) \geq \rho_0^*(DMU_q) \geq \rho_1^*(DMU_q)$.

The following Table 3 illustrates in detail the computation of the super-efficiency score for the unit DMU_5. First row of the table contains the original input and output values of this unit. The remaining rows (except row SBM) contain the virtual inputs and outputs given as follows:

$$\sum_{j=1}^{n} x_{ij}\lambda_j^*, \qquad i = 1, 2, ..., m$$

$$\sum_{j=1}^{n} y_{ij}\lambda_j^*, \qquad i = 1, 2, ..., r$$

where λ_j^* are optimum weights given by the super-efficiency models. The row SBM contains values x_i^* and y_i^* which are used for calculation of the super-efficiency measure in the model (5)-(6). The virtual units with the virtual inputs and outputs presented in Table 3 are always CCR efficient.

Table 3

Comparison of virtual units given by different models

Model	Inputs				Outputs	
	x^1	x^2	x^3	x^4	y^1	y^2
DMU_5	52.00	600.00	20.00	3.00	72.00	8.00
A-P	124.93	932.76	48.05	6.66	127.73	8.00
SBM	70.50	600.00	27.00	3.75	72.00	4.50
SBMG(t=0)	52.00	388.24	20.00	2.77	53.16	3.33
SBMG(t=1)	73.43	548.27	28.24	3.92	75.08	4.70

The super-efficiency score of the unit DMU_5 given by above used three different models is derived as follows (it is always a distance of the virtual unit and the evaluated unit):

1. *Andersen and Petersen model (3)*. The super-efficiency score $\theta^*(DMU_5)$ is derived as the maximum ratio of the virtual and the original inputs, ie. $\theta^*(DMU_5)$ = max[124.93/52, 932.76/600, 48.05/20, 6.66/3] = 2.4026. Due to the radial nature of this model the virtual unit on the efficient frontier lies often very far from the original unit and that is why the super-efficiency score can be very high. The conclusions following from this, the evaluated unit is efficient even its inputs increase $\theta^*(DMU_q)$-times, need not be always acceptable.

2. *Super SBM model (5)-(6)*. This model calculates the super-efficiency score as the ratio of two values. The numerator is the average expansion rate of inputs of the virtual unit comparing to the inputs of the evaluated unit and the denominator is the average reduction rate of outputs. The numerator in our example for the unit DMU_5 is (70.5/52 + 600/600 + 27/20 + + 3.75/3)/4 = 1.356. Similarly, the average reduction of outputs is (72/72 + 4.5/8)/2 = 0.781. Finally the super-efficiency score is $\delta^*(DMU_5)$ = = 1.356/0.781 = 1.5859. The virtual unit given by the SBM model is usually significantly closer to the original unit than in the previous model. That is why the super-efficiency score is here lower than the score given by the Andersen and Petersen model. It can be usually better explained and accepted for decision makers.
3. *Super SBMG model (9)*. The SBMG model minimises the sum of relative undesirable deviations (parameter t = 0) or the maximum relative deviation (t = 1). The undesirable deviations are positive slacks for inputs and the negative ones for outputs. The super-efficiency score in the first case is $\rho_0^*(DMU_5) = 1 + (72 - 53.16)/72 + (8 - 3.33)/8 = 1.8454$. Instead of the sum of deviations it could be possible to use their simple average. In this case the score equals to $\rho_0^*(DMU_5) = 1 + [(72 - 53.16)/72 + + (8 - 3.33)/8]/6 = 1.1409$. The minimisation of the maximum deviation leads to the optimum value $\rho_1^*(DMU_5) = 1 + (73.43 - 52)/52 = 1.4122$.

As the example shows, the ranking of the evaluated units defined by the presented super-efficiency characteristics is not always corresponding each other. Nevertheless this conclusion is typical for most multiple criteria decision making methods and corresponds to complexity of real decision problems.

CONCLUSIONS

The paper presents a new definition of super-efficiency in data envelopment analysis models. This new definition measures the distance of the evaluated real unit DMU_q and the virtual unit DMU* lying on the new efficient frontier by the positive and negative deviational variables that express the difference of the virtual inputs/outputs from the inputs/outputs of the unit DMU_q. The objective function of the super-efficiency model contains just the positive deviations for the inputs and the negative ones for the outputs. It is minimised in order to measure the distance in the undesirable way for both the groups of characteristics. Similarly to goal programming methodology either the sum of the undesirable deviation or the maximum deviation can be minimised. The proposed super-efficiency definition fulfils the basic requirement of the super-efficiency models, i.e. improving/worsening of any input or output has to lead to not worse/not better super-efficiency score. Our new definition has several advantages comparing to other ones:

- the super-efficiency model has always optimal solution, i.e. all the units receive their super-efficiency score,
- the model is non-radial – it works directly with the slacks of the inputs and outputs,
- the results of the model can be simply explained,
- the results of the model do not depend on its input or output orientation.

The results of the model were compared with other super-efficiency definitions on several numerical examples. All the comparisons show similarity of results in case the other models have a feasible solution. The future research will be concentrated on modification of the proposed model and its adaptation on specific conditions of the analysed sets of units.

REFERENCES

1. Andersen P., Petersen N.C.: A Procedure for Ranking Efficient Units in Data Envelopment Analysis. "Management Science" 1993, 39, pp. 1261-1264.
2. Charnes A., Cooper W.W., Lewin A., Seiford L.: Data Envelopment Analysis: Theory, Methodology and Applications. Kluwer Publ., 1994.
3. Cooper W.W., Seiford L.M., Tone K.: Data Envelopment Analysis. Kluwer Publ., 2000.
4. Jablonský J.: Super-Efficiency DEA Models and Their Applications. Proceedings of the 21st Conference Mathematical Methods in Economics, Czech Republic 2003, pp. 75-84.
5. Jablonský J.: Models for Efficiency Evaluation of Production Units. "Politická ekonomie" 2004, 52, pp. 206-220.
6. Tone K.: A Slack-Based Measure of Efficiency in Data Envelopment Analysis. "European Journal of Operational Research" 2001, 130, pp. 498-509.
7. Tone K.: A Slack-based Measure of Super-Efficiency in Data Envelopment Analysis. "European Journal of Operational Research" 2002, 143, pp. 32-41.
8. Zhu J.: Quantitative Models for Performance Evaluation and Benchmarking. Kluwer Publ., 2003.