Josef Jablonský

A SPREADSHEET-BASED SYSTEM FOR DEA MODELS[*]

Abstract

DEA models are tools for evaluation of efficiency and performance of decision making units. They are based on the definition of efficiency as the ratio of the sum of outputs produced by the unit divided by the sum of inputs spent in the production process. A standard LP solver is the only requirement for solving DEA models. Unfortunately specialized optimization packages are not available to typical users. In this case it is possible to use the built-in MS Excel solver. This solver has many limitations, but it is usually a sufficient tool for DEA models. The paper describes an original MS Excel add-on application that offers a simple tool for solving several standard DEA models. This application includes basic envelopment models with constant and variable returns to scale, SBM models, models with undesirable inputs and outputs and models with uncontrollable input and outputs. The application allows to calculate super-efficiency measures for most of the models mentioned. The functionality and main features of the system are illustrated by a simple case study – an evaluation of performance of pension funds in the Czech Republic.

Keywords

Data envelopment analysis, efficiency, Excel, spreadsheets.

INTRODUCTION

Data envelopment analysis (DEA) models are widely used as a tool for evaluation of efficiency, performance or productivity of homogenous decision making units, i.e. units that produce several identical or equivalent effects. These effects can be denoted as the outputs of the decision making units. We consider positive outputs of the unit, i.e. such that their higher values lead (assuming that other characteristics are unchanged) to higher performance

of the unit. To obtain the outputs the decision making units require several inputs that are usually minimized, i.e. their lower values lead to higher performance of the unit. Assuming the simplest case – one input and one output – the performance of the units can be simply expressed as the ratio:

$$\frac{\text{output}}{\text{input}}$$

In such a case we can receive many different financial characteristics with data that can be taken from financial statements of the evaluated unit. These simple ratio characteristics do not correspond to each other. That is why for the evaluation of the overall efficiency of the decision making unit it is necessary to take into account several inputs and outputs simultaneously.

Let us consider the set of homogenous units U_1, U_2, …, U_n described by r outputs and m inputs. Let us denote by $\mathbf{X} = \{x_{ij}, i = 1, 2, \ldots, m, \ j = 1, 2, ..., n\}$ the matrix of inputs and $\mathbf{Y} = \{y_{kj}, k = 1, 2, \ldots, r, \ j = 1, 2, ..., n\}$ the matrix of outputs. In general, the measure of efficiency of the unit U_q can be expressed as:

$$\frac{\text{weighted sum of outputs}}{\text{weighted sum of inputs}} = \frac{\sum_k u_k y_{kq}}{\sum_j v_j x_{jq}}$$

where v_j, $j = 1, 2, ..., m$ is the weight assigned to the j-th input and u_k, $k = 1, 2, ..., r$ is the weight of the k-th output. The evaluation of the efficiency of the unit U_q by a DEA model consists in maximization of its efficiency score under the constraints that the efficiency scores of all other units cannot be greater than 1 (100%). The weights of all inputs and outputs have to be greater than zero in order for the model to include all the characteristics. Such a model can be formulated as follows:

maximize $$\frac{\sum_k u_k y_{kq}}{\sum_j v_j x_{jq}}$$

subject to $$\frac{\sum_k u_k y_{kp}}{\sum_j v_j x_{jp}} \leq 1, \qquad p = 1,2,\ldots,n \tag{1}$$

$$u_i \geq \varepsilon, \qquad\qquad i = 1,2,\ldots,r$$

$$v_j \geq \varepsilon, \qquad\qquad j = 1,2,\ldots,m$$

The model (1) is known as a primal CCR (Charnes, Cooper, Rhodes) model. From the computational point of view it can be more efficient to work with the dual formulation:

minimize
$$z = \theta - \varepsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{k=1}^{r} s_k^+ \right)$$

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

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

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

where $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)$, $\lambda \geq 0$, is the vector of weights assigned to the evaluated units, $\mathbf{s}^+$ and $\mathbf{s}^-$ are vectors of positive and negative slacks in input and output constraints, ε is an infinitesimal constant and θ is a scalar variable expressing the reduction rate of inputs in order to reach the efficient frontier. The unit U_q is efficient if the following two conditions hold:

1. The optimum value of the variable θ^* is equal to 1.
2. The optimum values of all slacks $\mathbf{s}^+$ and $\mathbf{s}^-$ is equal to zero.

Apart from the information about the level of efficiency – the efficiency score θ^* – the DEA models compute inputs and outputs of the so-called virtual units. This unit lies always on the efficient frontier and expresses how to improve inputs/outputs of the evaluated unit in order to reach the efficient frontier. The virtual units corresponding to the units identified as efficient by a DEA model are identical because the efficient units lie on the frontier. The virtual units corresponding to non-efficient units can be expressed in the case of the model (2) as follows:

$$x_{iq}' = \theta^* x_{iq} - s_i^-, \qquad i = 1, 2, ..., m$$

$$y_{kq}' = y_{kq} + s_k^+, \qquad k = 1, 2, ..., r$$

The variables $\mathbf{s}^+$, $\mathbf{s}^-$ are exactly the slack variables expressing the difference between virtual inputs/outputs and the appropriate inputs/outputs of the unit U_q.

The CCR model (2) assumes constant returns to scale (CRS). The modification of the CCR model taking into account variable returns to scale (VRS) can be derived from the model (2) by adding the convexity constraint $\mathbf{e}^{\mathbf{T}}\boldsymbol{\lambda}=1$. Moreover, non-decreasing (NDRS) or non-increasing returns to scale (NIRS) can be considered by adding $\mathbf{e}^{\mathbf{T}}\boldsymbol{\lambda} < 1$ or $\mathbf{e}^{\mathbf{T}}\boldsymbol{\lambda} > 1$ respectively.

The model (2) is an input oriented DEA model, i.e. the aim of this model is to find how to reduce the inputs of non-efficient units in order to reach the efficient frontier. Similarly, it is possible to formulate an output oriented model. The basic modifications of the model (2) are given in the following list:

Input oriented models   **Output oriented models**   (3)

min
$$z = \theta - \varepsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{k=1}^{r} s_k^+ \right)$$
max
$$g = \phi + \varepsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{k=1}^{r} s_k^+ \right)$$

st
$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{iq}$$
st
$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{iq}$$

$$\sum_{j=1}^{n} \lambda_j y_{kj} - s_k^+ = y_{kq}$$

$$\sum_{j=1}^{n} \lambda_j y_{kj} - s_k^+ = \phi y_{kq}$$

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

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

CRS $\quad \sum_{j=1}^{n} \lambda_j - \text{free}$ CRS $\quad \sum_{j=1}^{n} \lambda_j - \text{free}$

VRS $\quad \sum_{j=1}^{n} \lambda_j = 1$ VRS $\quad \sum_{j=1}^{n} \lambda_j = 1$

NDRS $\quad \sum_{j=1}^{n} \lambda_j < 1$ NDRS $\quad \sum_{j=1}^{n} \lambda_j < 1$

NIRS $\quad \sum_{j=1}^{n} \lambda_j > 1$ NIRS $\quad \sum_{j=1}^{n} \lambda_j > 1$

The DEA models can be classified from many points of view. The aim of this paper is not to describe in detail the modifications of the above formulated model (2), but to discuss the possibility of solving DEA models in spreadsheets and describe our original application for their solving.

1. SOLVING DEA MODELS
IN SPREADSHEETS

The mathematical formulation of the DEA models shows that they can be solved as standard linear programming problems. The efficiency score for any of the decision making unit of the set of units is computed by solving one linear programming problem with $(n+m+r+1)$ variables and $(m+r)$ constraints. Even for a higher number of units (n) this is a low-sized LP problem that can be solved without difficulties by any of the professional optimization systems. In order to obtain the efficiency score for all the units the optimization problem of the mentioned size has to be solved n times. Problems with approximately one hundred units can be solved by means of any professional optimization systems in several seconds.

The built-in optimization solver in MS Excel is limited to problems with approx. 250 variables. This limit allows to solve DEA models (3) with $n = 200$ units and $m = r = 20$ inputs/outputs. The problem here is the necessity to repeat the optimization run n times in order to receive the appropriate results for all the units of the given set. That is why we decided to build an add-on application in the MS Excel environment that works with the internal MS Excel solver. In this way the system can be used on any computer with the MS Excels spreadsheet, i.e. on almost all computers. In this section we formulate the DEA models covered by the system and then describe how to work with this system.

The DEA Excel solver appears in the main Excel menu after its activation. The list of the available DEA models in the system is shown in Figure 1.
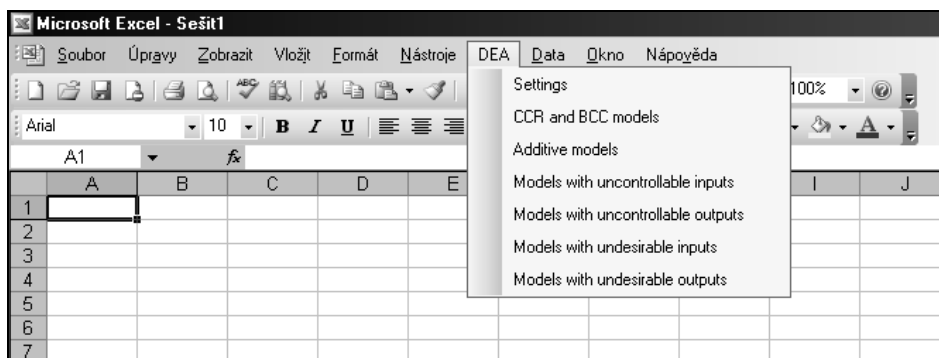

Fig. 1. The list of available DEA models

We will formulate them in detail in the following survey:

1. CCR and BCC models.

   They are often called envelopment models. These models were formulated in the introductory part of the paper – models (3).

2. Additive models.

   The additive models are often called SBM (slack based measure). This group of models measures the efficiency by means of slack variables only. In the application the following family of models in incorporated:

$$\text{max} \qquad z = \sum_{i=1}^{m} s_i^- + \sum_{k=1}^{r} s_k^+ \qquad\qquad (4)$$

$$\text{st} \qquad \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{iq}$$

$$\sum_{j=1}^{n} \lambda_j y_{kj} - s_k^+ = y_{kq}$$

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

$$\text{CRS} \qquad \sum_{j=1}^{n} \lambda_j - \text{free}$$

$$\text{VRS} \qquad \sum_{j=1}^{n} \lambda_j = 1$$

$$\text{NDRS} \qquad \sum_{j=1}^{n} \lambda_j < 1$$

$$\text{NIRS} \qquad \sum_{j=1}^{n} \lambda_j > 1$$

The objective function maximizes the sum of all slacks. Of course it is necessary to ensure the comparability of the inputs and outputs in this case. This can be simply done by normalization of all input and output values.

3. Models with uncontrollable inputs/outputs.

   In many applications some of the inputs or outputs cannot be directly controlled by the decision maker. In this case the uncontrollable characteristics have to be introduced into the model. The radial models (3) are modified as follows:

Input oriented models   **Output oriented models**   (5)

min
$$z = \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{k=1}^{r} s_k^+\right)$$

max
$$g = \phi + \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{k=1}^{r} s_k^+\right)$$

st
$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{iq} , \ i \in CI$$

st
$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{iq}$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{iq} , \ i \notin CI$$

$$\sum_{j=1}^{n} \lambda_j y_{kj} - s_k^+ = \phi y_{kq} , \ k \in CO$$

$$\sum_{j=1}^{n} \lambda_j y_{kj} - s_k^+ = y_{kq}$$

$$\sum_{j=1}^{n} \lambda_j y_{kj} - s_k^+ = y_{kq} , \ k \notin CO$$

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

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

CRS $\sum_{j=1}^{n} \lambda_j - \text{free}$ CRS $\sum_{j=1}^{n} \lambda_j - \text{free}$

VRS $\sum_{j=1}^{n} \lambda_j = 1$ VRS $\sum_{j=1}^{n} \lambda_j = 1$

NDRS $\sum_{j=1}^{n} \lambda_j < 1$ NDRS $\sum_{j=1}^{n} \lambda_j < 1$

NIRS $\sum_{j=1}^{n} \lambda_j > 1$ NIRS $\sum_{j=1}^{n} \lambda_j > 1$

where CI and CO is the set of indices of controllable inputs and outputs respectively. The family of models (5) formulated above is included in the DEA Excel solver.

4. Models with undesirable inputs/outputs.

 In typical cases inputs are to be minimized and outputs are to be maximized in DEA models, i.e. the lower value of inputs and the higher value of outputs lead to a higher efficiency score. It is not difficult to formulate a problem where some of the inputs and outputs will be of reverse nature. Such characteristics are denoted as undesirable inputs or outputs. Models with undesirable characteristics are included in the DEA Excel solver too. They are formulated as follows:

Input oriented (undesirable inputs)	**Output oriented (undesirable outputs)**	(6)

Input oriented (undesirable inputs):

min
$$z = \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{k=1}^{r} s_k^+\right)$$

st
$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{iq}, \; i \in DI$$

$$\sum_{j=1}^{n} \lambda_j x_{ij}' + s_i^- = \theta x_{iq}', \; i \notin DI$$

$$\sum_{j=1}^{n} \lambda_j y_{kj} - s_k^+ = y_{kq}$$

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

CRS $\quad \sum_{j=1}^{n} \lambda_j - \text{free}$

VRS $\quad \sum_{j=1}^{n} \lambda_j = 1$

NDRS $\quad \sum_{j=1}^{n} \lambda_j < 1$

NIRS $\quad \sum_{j=1}^{n} \lambda_j > 1$

Output oriented (undesirable outputs):

max
$$g = \phi + \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{k=1}^{r} s_k^+\right)$$

st
$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{iq}$$

$$\sum_{j=1}^{n} \lambda_j y_{kj} - s_k^+ = \phi y_{kq}, \; k \in DO$$

$$\sum_{j=1}^{n} \lambda_j y_{kj}' - s_k^+ = \phi y_{kq}', \; k \notin DO$$

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

CRS $\quad \sum_{j=1}^{n} \lambda_j - \text{free}$

VRS $\quad \sum_{j=1}^{n} \lambda_j = 1$

NDRS $\quad \sum_{j=1}^{n} \lambda_j < 1$

NIRS $\quad \sum_{j=1}^{n} \lambda_j > 1$

where DI and DO are the set of indices of desirable inputs and outputs respectively and $x'_{ij} = x^{max}_{ij} - x_{ij}, \, i \notin DI$, and $y'_{kj} = y^{max}_{kj} - y_{kj}, \, k \notin DO$.

The efficiency score in the abovementioned DEA models (except additive models) is limited to 1 (100%). Depending on the selection of the DEA model and on the relation between the number of units on the one hand and the number of inputs and outputs on the other hand, the number of efficient units can be relatively high. That is why several definitions of super-efficiency were formulated in order to classify the efficient units. In super-efficiency models the efficiency score of inefficient units remains unchanged (lower than 1 in input oriented models) and the efficiency score of efficient units is higher than 1. In this way the model makes it possible to classify the efficient units – this can be one of the very important results of the analysis. The original super--efficiency model is the Andersen and Petersen model [1]. This model adds to the models (2), (3), (5), and (6) a new constraint $\lambda_q = 0$ (the weight of the

evaluated unit is constant and equal to zero). This super-efficiency option can be used in the DEA Excel solver by checking the appropriate box. Nevertheless, it is necessary to mention that a feasible (and optimum) solution of the super--efficiency models exists always exactly under the assumption of constant returns to scale.

2. USING DEA EXCEL SOLVER – A SIMPLE CASE STUDY

The work with the system will be demonstrated on a small numerical example with a real economic background. It is the problem of evaluating the efficiency of the available pension funds in the Czech Republic. We have worked with the data set for 12 pension funds, each of them characterized by the following seven criteria (the data are from the year 2003):
1. INP 1 – the number of customers [thousands],
2. INP 2 – total assets [mil. CZK],
3. INP 3 – equity capital [mil. CZK],
4. INP 4 – total costs [mil. CZK],
5. OUT 1 – appreciation of the customer deposits for the last year (2003) [%],
6. OUT 2 – average appreciation of the customer deposits for the last three years (2001-2003) [%],
7. OUT 3 – net profit [mil. CZK].

For the DEA analysis, the first four criteria were taken as inputs and the remaining ones as outputs of the model. The data set for evaluation is given in the spreadsheet in Figure 2.

	#of cust.	assets	equity	tot. costs	appr. 1	appr. 3	profit
Allianz	106	4095	77	49,5	3	3,69	1,29
Credit Suisse	611	22592	549	454,1	3,36	3,67	5,22
CSOB Progres	18	452	56	15,1	4,3	4,15	1,13
CSOB Stabilita	304	8508	298,6	203,3	2,3	2,83	10,87
Generali	23	789	74	15,5	3	3,9	0,45
ING PF	346	9767	289,1	221,7	4	4,27	0,26
CP PF I	225	6348	290,7	184,7	3,34	3,65	6,83
CP PF II	518	12441	522,5	297,3	3,1	3,37	6,9
CS PF	401	10954	223,5	238,8	2,64	3,31	1,1
KB PF	285	11776	441,6	166	3,4	4,14	6,4
PF Ostrava	19	935	71	18,2	2,44	2,68	0,04
PF Zemsky	14	468	87,9	23,2	4,01	4,24	2,03

Fig. 2. Evaluation of pension funds – the data set

The new DEA menu item in MS Excel contains just the *Settings* option and then the selection of the DEA model that will be used for analysis (Figure 1). The *Settings* item contains the possibility to specify several parameters of the system, but they need not be changed as they are set to their defaults:

- *Language* – one of the two available language versions of the system (English, Czech),
- *Tolerance* – a constant with the initial value 10^{-6} which is used for testing of zero variables values (MS Excel solver often returns values very close to zero instead of zeroes),
- *Title* – the text displayed in the header part of the output of results,
- *Epsilon* – a constant of infinitesimal value – an initial value is 10^{-8},
- *Normalization of input data* – a switch (on/off) with the initial value "on" which specifies whether the normalization of input data should be conducted (a transformation to a comparable scale).

Let us suppose that the decision maker wants to apply the radial model with variable returns to scale (BCC model). After the appropriate family of models is selected from the main menu – *CCR and BCC models* in this case – the dialog box appears and the user can insert all the necessary information. The dialog box contains the following items:

- *DMU's labels* – a range with labels of the evaluated units (not obligatory – when it is not specified the system works with the default labels DMU1, DMU2,...),
- *Input/output labels* – ranges with the labels of the inputs and outputs (not obligatory – when it is not specified the system works with the default labels INP1, INP2,... and OUT1, OUT2...),
- *Matrix of inputs and outputs* – two continuous ranges containing the matrix of inputs and the matrix of outputs – in our example it is the range B2:E13 for the inputs and F2:H13 for the outputs,
- *Model orientation* – one of the two choices: input- or output-oriented model,
- *Returns to scale* – one of the four choices: CRS, VRS, NIRS, NDRS,
- *Super-efficiency* – the switch that sets up the selection of the super--efficiency model,
- *Optimization in two steps* – the switch that specifies whether the optimization is realized in one or two steps (the first step is the optimization of the reduction variable θ or the expansion variable ϕ and the second is the maximization of the slack variables $\mathbf{s}^+$ and $\mathbf{s}^-$,

– *Detailed/brief output of results* – two choices that switch on/off a brief and/or detailed output of the results – for the results the system creates single sheets with output information.
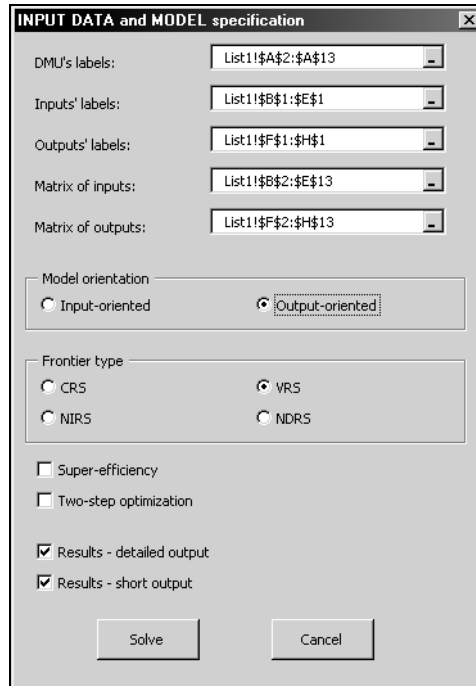

Fig. 3. Input data and model selection

The brief output information sheet for our example is presented in Figure 4. The sheet contains the following information:
1. The specification of the DEA model used in the analysis (VRS_O is an output oriented model with variable returns to scale).
2. The DMU labels and the efficiency scores computed by the model (the efficient units are marked by red colour).
3. The values of virtual inputs and outputs (target values for reaching the efficient frontier).
4. Non-zero weights of the units (the linear combination of units using these weights gives the virtual inputs and outputs).
Except this information the detailed output sheet contains the optimum values of slack variables and the original values of inputs and outputs.

The results of the BCC model for our example are given in Figure 4. It is clear that among twelve pension funds six were identified as efficient by the selected DEA model. The "worst" fund is PF Ostrava that has to improve its outputs by more than 56% in order to reach the efficient frontier. The brief output information sheet contains target values, i.e. the values of the input and output characteristics for improving the efficiency and reaching the efficient frontier.

Fig. 4. A brief output information sheet

CONCLUSIONS

The DEA Excel Solver described in the previous sections can be downloaded from the download section of the web page http://nb.vse.cz/~jablon and used by any interested professionals. The application can solve problems up to 200 decision making units and 20 inputs and the same number of outputs. This size is sufficient for most of the real-world problems. The application will be extended in the future by other DEA models. The advantage of the system is that it does not assume any specialized software products except MS Excel including the built-in Excel optimization solver that is available on almost all computers. The functionality of the system was illustrated on a simple case study – the evaluation of the efficiency of pension funds in the Czech Republic. Even though this study contains only twelve decision making units, the system can solve any other problem (of limited size) within several seconds.

REFERENCES

1. Andersen P. and Petersen N.C.: A Procedure for Ranking Efficient Units in Data Envelopment Analysis. "Management Science" 1993, 39, pp. 1261-1264.
2. Cooper W.W., Seiford L.M., and Tone K. (2000): Data Envelopment Analysis. Kluwer Academic Publishers, 2000.
3. Jablonský J.: Measuring Efficiency of Production Units by AHP Models. "Mathematical and Computer Modeling" 2007, 46.
4. Jablonský J.: A MS Excel Based Support System for Data Envelopment Analysis Models. In. Ed. H. Skalská. Proceedings of the 23rd Conference Mathematical Methods in Economics. Hradec Králové 2005, pp. 175-181.
5. Jablonský J.: Models for Efficiency Evaluation of Production Units. "Politická Ekonomie" 2004, 52, pp. 206-220.
6. Zhu J.: Quantitative Models for Performance Evaluation and Benchmarking. Kluwer Academic Publishers, 2003.