DEA by sequential exclusion of alternatives

Fuad Aleskerov (alesk@hse.ru) [1,2]

Vsevolod Petrushchenko (goroddt@yandex.ru) [1]

Abstract

Data Envelopment Analysis is a well-known non-parametric technique of efficiency evaluation which is actively used in many economic applications. However, DEA is not very well applicable when a sample consists of firms operating under drastically different conditions. We offer a new method of efficiency estimation on heterogeneous samples based on a sequential exclusion of alternatives and standard DEA approach. We show a connection between efficiency scores obtained via standard DEA model and the ones obtained via our algorithm. We also illustrate our model by evaluating 28 Russian universities and compare the results obtained by two techniques.

Key words: efficiency, Data Envelopment Analysis, sequential exclusion of alternatives, universities' efficiency.

[1] *National Research University Higher School of Economics (NRU HSE), Moscow, Russian Federation*

[2] *Institute of Control Science, Russian Academy of Science (ICS RAS), Moscow, Russian Federation*

1 Introduction

In standard DEA approach many firms can be located quite far from efficiency frontiers. From the economic point of view it means that all inefficient firms are benchmarked against some outstanding companies which are very rare in the whole sample. However, there are a lot of other much more complicated situations in which different types of heterogeneity make it impossible to use standard DEA technique successfully.

In our paper, we concentrate mainly on the heterogeneity caused by drastic differences in operating environment. These are (for details see Fried et. al., (1999)

1. Differences in ownership status (public/private, corporate/non-corporate);
2. Location peculiarities (for universities — city/country, for electrical companies — the density of population in the operating area);
3. Differences in legislation.

There are a lot of papers dealing with the problem of the influence of environmental parameters on efficiency scores, see Banker and Morey (1986a, 1986b), Charnes, Cooper and Rhodes (1981), Bessent and Bessent (1980), Ferrier and Lovell (1990). The following solutions are commonly used

1. Partitioning of an original sample to the smaller groups by some environmental factor (for instance, location in city – first group, suburbs – second group, etc). Comparison is performed only between subsamples;
2. Separate application of DEA to each cluster, then construction of each firm's projection onto its respective efficiency frontier and launching one common DEA LP among the obtained projections;
3. Imposition of additional restrictions to the DEA;
4. Composition of regression analysis with the approach 3.

One can find the detailed description of all methods, their strengths and shortages in Coelli et al (2005). There is another widespread technique of taking into account heterogeneity of the sample. The idea is to combine the power of clustering models with DEA (see, e.g., Samoilenko, K.M.Osei-Bryson (2010); Shin and Sohn, (2004); Hirschberg and Lye, (2001); Lemos et al., (2005); Meimand et al., (2002); Sharma and Yu, (2009); Marroquin et al., (2008); Schreyogg and von Reitzenstein, (2008)). Generally, clustering methods can be united with DEA via two different ways. The first one is to apply clustering to the obtained efficiency scores, then form appropriate reference subsets of firms and apply DEA again. The second one is on the contrary based on the application of clustering to initial set of DMUs and then comparison of each firm within its reference set.

Another approach introduced by Smirlis and Despotis, (2012) deals with the problem of extreme units in DEA. The firm is characterized as extreme if its level of output (input) is too high (low) for the given sample. First, the authors formulate linear program which allows to reduce the influence of such extreme firms on efficiency scores. Second, they offer a technique for specifying a threshold beyond which the values of output/input are characterized as extreme.

Aggregated ratio analysis is another model introduced by Huang et al., (2005). Authors claim their model is equivalent to standard CCR. However, Zha and Liang (2014) found some errors in the proof of Theorem 1.

The so-called context dependent DEA introduced by Chen et al., (2005) is another model aimed at heterogeneity problem in efficiency assessment. Authors state that the model measures the relative attractiveness of evaluated items on a specific performance level against items exhibiting poorer performance.

The structure of our algorithm differs from the above mentioned approaches. We suggest to move the efficiency frontier in a special way, using the barycenter of the sample.

The next section presents the method in the simplest possible case with economy consisting of single input and output. Then we introduce one of the possible ways to extend our model to the evaluation of samples with arbitrary number of inputs and outputs. In the last section we illustrate the proposed method via evaluating the efficiency scores of 28 Russian universities and comparing them with standard DEA.

2 The model

First, we briefly discuss Data Envelopment Analysis, which is one of the most widespread and commonly used techniques of efficiency evaluation. The method offered by Farrell (1957) was extended and generalized by Cooper et al (1978). They showed that the problem of efficiency evaluation can be formulated in terms of mathematical program as

$$\max_{u,v}\left(\theta_i = \frac{u_1 q_{1i} + \ldots + u_M q_{Mi}}{v_1 x_{1i} + \ldots + v_N x_{Ni}}\right)$$

subject to

$$\begin{cases} \frac{u_1 q_{1i} + \ldots + u_M q_{Mi}}{v_1 x_{1i} + \ldots + v_N x_{Ni}} \leq 1, i \in \{1, \ldots, L\}; \\ u_j \geq 0, j \in \{1, \ldots, M\}; \\ v_k \geq 0, k \in \{1, \ldots, N\}, \end{cases}$$

where L is the number of firms in the sample, q_{ji} — j-th output parameter ($j \in \{1, \ldots, M\}$) of i-th firm, x_{ki} — k-th input parameter ($k \in \{1, \ldots, N\}$) of i-th firm, u and v are weight vectors of appropriate lengths. Finally, θ_i represents efficiency measure of i-th firm.

Recall the standard definition of DEA using CRS setting. Cooper et al (1978) also showed that presented above model can be simplified and rewritten in the form of linear program as

$$\min_{\lambda, \theta_i} \theta_i \tag{1}$$

subject to

$$\begin{cases} -q_i + Q\lambda \geq 0; \\ \theta_i x_i - X\lambda \geq 0; \\ \lambda \geq 0, \end{cases} \tag{2}$$

where q_i is $M \times 1$ vector of output parameters of i-th firm, x_i is $N \times 1$ vector of input parameters of i-th firm, Q is $M \times L$ matrix of output parameters of all firms, X is $N \times L$ matrix of input parameters of all firms, λ is $L \times 1$ weight vector, one may interpret it as *intensity parameters*, (Coelli, 2005). As in the previous case θ_i is the efficiency measure of i-th firm.

The formulation (1)–(2) is fundamental and called CCR model (after the names of its authors). Note that CCR allows to assess efficiency only when constant return to scale takes place. Thereby the program (1)–(2) is also called CRS DEA model. However, it is easy to adjust the method to the situation with variable return to scale – we only need to impose one additional constraint

$$\mathbf{1}^T \cdot \lambda = 1, \tag{3}$$

where $\mathbf{1}^T$ is a unit vector of the size $1 \times L$.

The programm (1)–(2) with the restriction (3) is called VRS DEA model. This modification was introduced in Charnes and Cooper (1984). Note that VRS model is applicable if analyzed firms operate at the non-optimal scale.

One can write the linear programm dual to (1), (2), (3) and discover a geometric interpretation of the two discussed models in the case of single input and output production (*Fig.* 1).

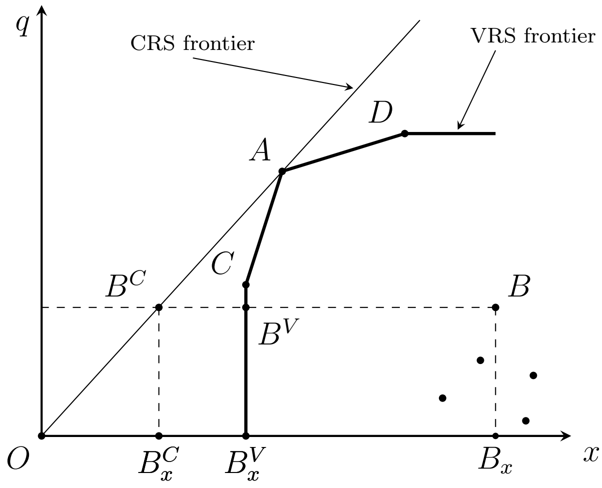

Figure 1. Interpretation of two DEA models in the case $N = M = 1$.

Efficiency score of the firm B (*Fig.* 1) via VRS and CRS models is calculated as $\frac{|OB_x^V|}{|OB_x|}$ and $\frac{|OB_x^C|}{|OB_x|}$, respectively.

Throughout the rest of the text we use the definition of efficient firms according to CRS model. In this section we consider the situation with single input and output. Note that in this case there may be several efficient firms if and only if all of them are lying on the same ray which i) begins in the origin, and ii) has the highest slope amongst all analogous rays which connect other firms with origin. Algebraically it means that several efficient firms must have exactly the same minimal among others ratio of input to output. Therefore without loss of generality we consider the case when there is only one 100% efficient company in the sample.

Our purpose is to construct a new efficiency frontier which takes into account heterogeneity of the evaluated sample. Recall that the i-th firm in the sample is represented via two coordinates (x_i, q_i) in the space of input-output parameters.

The core idea is the following. First, we calculate the barycenter of all firms in the usual geometric sense. The next step is to construct a frontier generating company lying on the segment between the most efficient firm and the barycenter of the sample (*Fig.* 2). Now it is possible to form a subgroup of relatively inefficient organizations and evaluate their scores regarding the new frontier generating company.

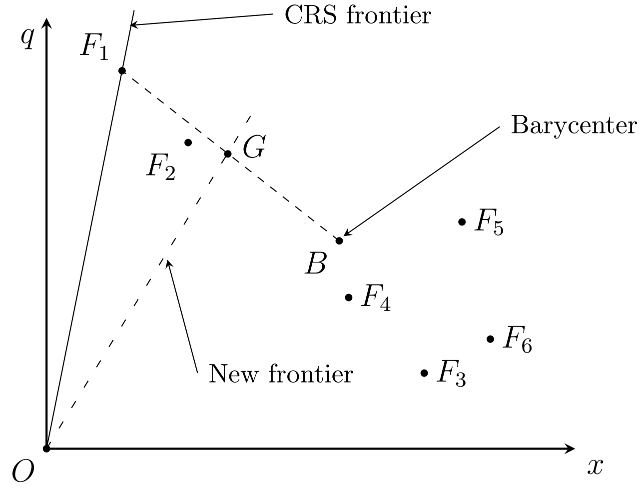

Figure 2. Graphic interpretation of the algorithm in the case $N = M = 1$.

On the figure above initial sample consists of the firms $F_1, \ldots, F_6$ and according to standard CRS model F_1 is the efficient firm. According to the introduced algorithm we calculate the barycenter (point B) and construct the new frontier via generating phantom firm G, lying on the segment BF_1. Clearly, $F_3, \ldots, F_6$ should be benchmarked against the firm G. Still, F_1 and F_2 remain unevaluated, to compute their efficiency scores we should repeat the same algorithm excluding the firms $F_3, \ldots, F_6$ from consideration.

It is of separate interest to define exact position of the frontier generating company. It is clear that this location depends on the heterogeneity of the sample. Roughly speaking, it means that the higher heterogeneity within the sample the nearer generator to the barycenter. Let us measure heterogeneity of a sample as a number $\mu \in [0, 1]$. The higher heterogeneity the higher the value of μ. It is out of scope of this paper to discuss how one can compute this index. Then the position of generating company is defined as

$$G = \mu B + (1 - \mu) F_1, \tag{4}$$

where B is the barycenter of the sample and F_1 is a 100% efficient firm according to the standard DEA CRS model.

Note several important properties of the procedure. First, the algorithm obviously converges for any sample. Second, the only firm that remains efficient is the one which is efficient according to standard CRS model. Besides there is a simple connection between DEA efficiency scores and the ones obtained via the sequential process. Suppose that current subgroup of firms is evaluated via frontier generating firm G. Let F be in this subgroup, then

$$E_F^{CRS} = E_G^{CRS} \cdot E_F^{New}, \tag{5}$$

where the lower index stands for firms and the upper one does for efficiency evaluation method. Note that the formula (5) follows immediately from the interpretation of CRS efficiency scores given in *Fig. 1*.

According to (5) our algorithm evaluates inefficient firms less strictly than the standard CRS model. Again, the reason for such alleviation is that the sample is heterogeneous and all firms cannot be benchmarked against the firm which showed exceptional efficiency. Such situations happen in practice very often and may occur, for instance, because of the presence of some crucial *environmental* factors.

3 Extensions to general case

Consider now the case of a sample characterized by several input and output variables. It is impossible to apply the considerations above directly. Thus we construct a sequence of linear programs which allow us to carry out the same algorithm in general case. Besides, we want to preserve the following properties

i) Convergence of the procedure for any sample;

ii) The only firms that remain efficient are those which were efficient according to standard CRS model;

iii) Some counterpart of the equality (5) should be obtained.

Recall that i-th firm is represented by the vector $x_i = (x_{1i}, \dots, x_{Ni})$ of inputs and $q_i = (q_{1i}, \dots, q_{Mi})$ of outputs. As before we define the input and output parts of the barycenter as

$$b_x = (\overline{x}_1, \dots, \overline{x}_N).$$

and

$$b_q = (\overline{q}_1, \dots, \overline{q}_M),$$

where, as usual, the bar means the average value of a particular parameter.

We will also need the heterogeneity index μ.

Let the whole sample be defined by the set of indices $I = \{1, \dots, L\}$. Recall that there are N input and M output variables. We denote the group of 100% efficient (relatively to the standard DEA) companies as a subset $I_e = \{i_1, \dots, i_S\} \subset I$. Now, let X_e be the $N \times S$ matrix of input parameters of all efficient firms, and Q_e be the $M \times S$ matrix of outputs for the same firms. We also define the following matrix

$$B_x^i = \|\underbrace{b_x^T, \dots, b_x^T}_{S}, x_i^T\|,$$

where b_x^T is transposed input part of the barycenter repeated S times, x_i is the input vector for some inefficient firm, i.e., $i \in I \setminus I_e$. Similarly we define the $M \times (S+1)$ matrix

$$B_q^i = \|\underbrace{b_q^T, \dots, b_q^T}_{S}, q_i^T\|,$$

where b_q^T is transposed output part of the barycenter, and q_i is the vector of outputs for the i-th company, $i \in I \setminus I_e$. Let X_e^i and Q_e^i be the matrices X_e and Q_e with the one added column — x_i^T and q_i^T, respectively. Since the core idea is to move the frontier towards the barycenter, we can form the matrices

$$X_i = \mu B_x^i + (1-\mu) X_e^i \text{ and } Q_i = \mu B_q^i + (1-\mu) Q_e^i, \tag{6}$$

where the product of a matrix by a scalar is defined in the usual componentwise way.

Let us make two important remarks. First, matrices (6) are defined only for inefficient companies, i.e. $i \in I \setminus I_e$. Note that the last column of X_i and Q_i are x_i^T and q_i^T, respectively.

Now we introduce the general form of the procedure. The first step is to solve the following linear program for every inefficient firm $i \in I \setminus I_e$.

$$\min_{\lambda, \theta_i^*} \theta_i^* \tag{7}$$

subject to

$$\begin{cases} -q_i + Q_i \lambda \geq 0; \\ \theta_i^* x_i - X_i \lambda \geq 0; \\ \lambda \geq 0, \end{cases} \tag{8}$$

where X_i and Q_i are defined in (6), λ is $(S+1) \times 1$ vector of constants, x_i and q_i are input and output vectors for the i-th inefficient firm. Finally, θ_i^* is the corrected efficiency score of the i-th inefficient company.

The algorithm works as follows. Let $Z_1 = I \setminus I_e$, i.e., Z_1 is the set of all firms inefficient according to the standard DEA model. The description of k-th stage is as follows

1. Calculation of the barycenter of firms included in the set $Z_k \cup I_e$;

2. Calculation of new X_i and Q_i matrices for all companies i in Z_k;

3. Calculation of $\theta_i^* < 1$ for all companies i in Z_k;

4. All those companies i which get $\theta_i^* < 1$ are excluded from the sample, i.e. $Z_{k+1} = Z_k \setminus \{i|\theta_i^* < 1\}$;

5. If Z_{k+1} is empty then stop, if not – begin $(k+1)$-th stage of the algorithm.

We did not take into account only one case, when matrices (6) are organized in such a way that for all inefficient companies according to (7)-(8) $\theta^* = 1$ holds, i.e., $Z_{k+1} = Z_k$ at some stage k. It means that the original frontier is moved too much. Therefore we have to decrease the value of μ and begin the procedure from the beginning. For instance, we can take the new value of the heterogeneity index as μ^2.

To conclude we make two remarks regarding the algorithm. The convergence is guaranteed by construction. The set of efficient firms is preserved as well. Although we cannot preserve the property (5), the straightforward counterpart is the following. Since we use standard DEA CRS model, we can calculate a projection of every inefficient firm on the temporary frontier defined by (7)-(8) programm (see Coelli (2005) for details). Then it holds that

$$E_F^{CRS} = E_P^{CRS} \cdot E_F^{New}, \tag{9}$$

where E_F^{CRS} is the standard efficiency score of the firm F, E_P^{CRS} is the efficiency of the projection P of the firm F on the new frontier defined by (7)-(8). Finally, E_F^{New} is the efficiency score of the firm F according to our procedure.

Thus, we proposed a theoretical description of the sequential DEA process and showed the simplest properties of the procedure. Note also that our model with $\mu = 0$ corresponds to the usual DEA CRS model.

Now, let us give an example how the algorithm works for the case of two inputs and single output parameter.

As always, we denote first input, second input, single output and heterogeneity index defined in (4) as x_1, x_2, q and μ, respectively. Apparently all efficient firms with coordinates (x_1^e, x_2^e, q^e) are moved towards the barycenter via the following transformation

$$(x_1^e, x_2^e, q^e) \longmapsto \mu(\bar{x}_1, \bar{x}_2, \bar{q}) + (1-\mu)(x_1^e, x_2^e, q^e). \tag{10}$$

It is known that in this case standard CRS model can be easily visualized on the plane $(\frac{x_1}{q}, \frac{x_2}{q})$. Without loss of generality, suppose all outputs are equal to 1, it allows to simplify the plane to the form (x_1, x_2). Recall also that $\mu\bar{q} + (1-\mu)q^e = 1$. Therefore, according to (10) every frontier generating company has the following coordinates on the space (x_1, x_2)

$$(\mu\bar{x}_1 + (1-\mu)x_1^e, \mu\bar{x}_2 + (1-\mu)x_2^e)\,.$$

This means that frontier generating companies are represented as convex combination of the barycenter and efficient according to standard CRS model firms. Generally (if outputs are arbitrary) it is not the case, however, it adds only technical difficulties.

Using all above notes we are able to illustrate the first phase of the algorithm on *Fig. 3*.

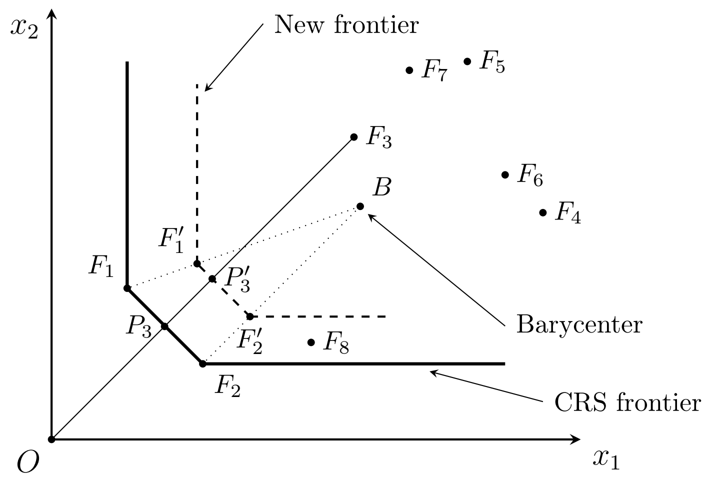

Figure 3. Graphic interpretation of the algorithm for the case $N = 2, M = 1$.

According to *Fig. 3*, F_1 and F_2 are efficient according to the standard DEA model, therefore we move them towards the barycenter (with the fixed μ). Thus, F_1' and F_2' become new frontier generating firms. It is clear that on the first stage of the procedure we evaluate only $F_3, \ldots, F_7$ because F_8 gets $\theta_8^* = 1$. For instance, the efficiency score of F_3 can be determined as a ratio $|OP_3'|$ to $|OF_3|$, where P_3' is the projection of F_3 on the new frontier. Note that standard efficiency of this firm is defined by $\frac{|OP_3|}{|OF_3|}$, where P_3 is a similar projection of F_3 onto standard CRS frontier.

Finally, let us illustrate what the identity (9) means

$$E_{P_3'}^{CRS} \cdot E_{F_3}^{New} = \frac{|OP_3|}{|OP_3'|} \cdot \frac{|OP_3'|}{|OF_3|} = \frac{|OP_3|}{|OF_3|} = E_{F_3}^{CRS},$$

where all notation is taken from *Fig. 3*.

To conclude this Section, let us note also that the procedure cannot be simplified and performed via some single-step modification of the standard DEA model.

4 An illustration

We apply now our model to evaluation of efficiency scores for 28 Russian universities and compare the results with the standard DEA outcome. The detailed description of related research is presented in Abankina et. al. (2012). To apply DEA we choose three input parameters, which reflect main universities' resources – the level of state financing, quality of professorial and teaching staff and a quality of entrants.

1. Funding from federal budget (denoted as I_1);

2. The number of employees with a degree of Doctor of Science (denoted as I_2);

3. The quality of university entrants, to estimate this parameter we use a mean value of Universal State Exam (USE), which is mandatory for admission (denoted as I_3).

and two output parameters

1. The number of students who do not pay tuition (denoted as Q_1);

2. The number of published articles on refereed journals (denoted as Q_2).

The first output indicates the attractiveness of a university for the applicants and the second one is a proxy for success of scientific and research work within a university. The descriptive statistics for all parameters is presented below (28 observations for 2008).

Table 1. Descriptive statistics of input and output parameters

	I_1	I_2	I_3	Q_1	Q_2
Mean value	416486.71	68.35	454.39	424.35	612.03
Variance	156141404288.75	45.30	553190.76	56397.94	675861.36
Standard Deviation	395147.31	6.73	743.76	237.48	822.10
Median	261386.5	69.05	204	362.5	315
Minimum	61190.9	56.7	18	116	6
Maximum	1694875.5	82.6	3770	926	3556
Sum	11661628.1	1913.7	12723	11882	17137

First, we calculate efficiency scores according to standard DEA model. After that we apply our technique taking three distinct values of heterogeneity index μ, namely, 0.2, 0.5 and 0.8. We emphasize that these values are chosen only to test the developed model, we did not computed the index of heterogeneity of Russian universities in a fair way. Appendix 1 contains the detailed list of efficiency scores for all four cases.

We compare the results obtained via different models in two ways. First, we rank all universities according to their efficiency scores in each of four cases and compare different orderings via Kendall's distance, see Kendall (1938), i.e., we count all discordant pairs in the two ranks and then normalize this value by dividing by the total number of pairs in a list consisting of N objects. The discordant pair (i, j) is the one for which i is better than j in the first rank and j is better than i in the second one, or vice versa. Consequently a concordant pair is the one which is ranked in the same order in both orderings.

Further, let us denote the number of discordant pairs as N^- and the number of concordant pairs as N^+. Note that

$$N^+ + N^- = C_N^2 = \frac{N(N-1)}{2},$$

where C_N^2 is a binomial coefficient.

According to this notation the Kendall's distance may be calculated as

$$K(r_1, r_2) = \frac{N^+ - N^-}{N^+ + N^-}, \tag{11}$$

where r_1 and r_2 are different ranks consisting of N objects. Note that the value of Kendall's distance lie between -1 and 1, where 1 means that two orderings are the same and -1 means that the two rankings are inverse.

Table 2 shows the Kendall's distance between all four types of efficiency evaluation models.

Table 2. Kendall's distances

	DEA	$\mu = 0.2$	$\mu = 0.5$	$\mu = 0.8$
DEA	1	-	-	-
$\mu = 0.2$	0.9086	1	-	-
$\mu = 0.5$	0.7849	0.8441	1	-
$\mu = 0.8$	0.7258	0.7097	0.6720	1

Note that the distance between ranks obtained via the technique is small. Moreover the nearer the value of μ to 1 the higher the bias of new efficiency scores from the ones obtained via standard DEA.

As another measure of a difference between two rankings we compute median, mean and minimal values of efficiency scores for all four versions of our evaluations (recall that standard DEA can be obtained from our model taking the value of $\mu = 0$). The information is given below.

Table 3. Median and mean values of efficiency scores for different models (in percents)

	DEA	$\mu = 0.2$	$\mu = 0.5$	$\mu = 0.8$
Median	53.19	61.04	88.98	92.35
Mean	61.60	68.94	81.09	86.56
Minimal	15.25	22.16	26.64	32.16

Again, Table 3 confirms that the obtained results are quite consistent. With the increasing of heterogeneity index μ our model evaluate all firms more and more mildly. All characteristics are increasing with the growth of μ.

5 Conclusion

We have introduced a new algorithm of efficiency evaluation in the case when the sample is heterogeneous. One of the main assumptions used is that the geometric barycenter of a sample represents the average situation of the evaluated sector of economy. Taking it into account, the core idea of our technique is to move the efficiency frontier towards the barycenter. It allows to evaluate all inefficient firms more mildly.

Our algorithm has three important properties. First, the convergence for any sample is guaranteed. Second, the set of firms that are efficient according to the standard DEA model is preserved. Finally, there is the simple connection between efficiency scores obtained via DEA and our algorithm.

We tested our model on the real data set containing the information on five parameters on 28 Russian universities in 2008. The developed technique shows consistent results, i.e., our model does not crucially change the structure of a ranking by efficiency, however, the efficiency scores grow when the heterogeneity of the sample is increasing.

6 Acknowledgement

We thank Decision Choice and Analysis Laboratory and Scientific Fund of HSE for partial financial support.

7 Appendix 1

Efficiency scores (in percents) for different models

№	DEA ($\mu = 0$)	$\mu = 0.2$	$\mu = 0.5$	$\mu = 0.8$
1	100	100	100	100
2	54	60	74	98
3	100	100	100	100
4	35	40	77	91
5	42	59	90	78
6	53	61	96	93
7	99	99	99	99
8	15	35	38	32
9	24	27	40	71
10	100	100	100	100
11	23	26	33	58
12	31	39	73	81
13	53	59	71	90
14	70	98	99	85
15	100	100	100	100
16	100	100	100	100
17	100	100	100	100
18	73	84	98	90
19	19	22	26	33
20	100	100	100	100
21	45	52	57	74
22	52	58	84	81
23	45	52	79	88
24	59	80	95	89
25	49	55	67	93
26	92	98	94	97
27	33	47	87	97
28	48	68	82	97
Mean	61.60	68.94	81.09	86.56

8 References

1. Abankina, I.V., Aleskerov, F.T., Belousova, V.Y., Bonch-Osmolovskaya, A.A., Petrushchenko, V.V., Ogorodniychuk, D.L., Yakuba, V.I., Zin'kovsky K.V., (2012), "University Efficiency Evaluation with Using its Reputational Component", *Lecture Notes in Management Science*, Tadbir Operational Research Group, Ltd, 4, 244-253.

2. Banker, R.D., A. Charnes and W.W. Cooper (1984), "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis", *Management Science*, 30, 1078-1092.

3. Banker, R.D., and R.C. Morey (1986a), "Efficiency Analysis for Exogenously Fixed Inputs and Outputs", *Operations Research*, 34, 513-521.

4. Banker, R.D., and R.C. Morey (1986b), "The Use of Categorical Variables in Data Envelopment Analysis", *Management Science*, 32, 1613-1627.

5. Bessent, A.M., and E.W. Bessent (1980), "Comparing the Comparative Efficiency of Schools through Data Envelopment Analysis", *Educational Administration Quarterly*, 16, 57-75.

6. Charnes, A., W.W. Cooper and E. Rhodes (1978), "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research*, 2, 429-444.

7. Charnes, A., W.W. Cooper and E. Rhodes (1981), "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through", *Management Science*, 27, 668-697.

8. Chen Y., Morita H., Zhu J. (2005), "Context-dependent DEA with an application to Tokyo public libraries", *International Journal of Information Technology and Decision Making*, 4, 385-394.

9. Coelli, T.J., D.S.P. Rao, C.J. O'Donnell, G.E. Battese (2005), "An Introduction to Efficiency and Productivity Analysis", 2nd edition, Springer, N.Y..

10. Despotis D.K. , Smirlis Y.G. (2012), "Relaxing the impact of extreme units in Data Envelopment Analysis", *International Journal of Information Technology and Decision Making*, 11, 893-907.

11. Farrell, M.J. (1957), "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society, Series A*, CXX, Part 3, 253-290.

12. Ferrier, G.D., and C.A.K. Lovell (1990), "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence", *Journal of Econometrics*, 46, 229-245.

13. Fried, H.O., S.S. Schmidt and S. Yaisawamg (1999), "Incorporating the Operating Environment into a Nonparametric Measure of Technical Efficiency", *Journal of Productivity Analysis*, 12, 249-267.

14. Hirschberg, J.G., J.N. Lye, (2001), "Clustering in a Data Envelopment Analysis Using Bootstrapped Efficiency Scores", *Department of Economics – Working Papers Series 800*, The University of Melbourne.

15. Huang, Z., Liang, L., Susan X. Li, D. Wu, (2005), "Aggregated ratio analysis in DEA", *International Journal of Information Technology and Decision Making*, 4, 369-384.

16. Kendall, M.A., (1938) "New Measure of Rank Correlation", *Biometrika*, 30, 81–89.

17. Lemos, C.A.A., M.P. Lima, N.F.F. Ebecken, (2005), "DEA Implementation and Clustering Analysis using the K-means algorithm", *Data Mining VI – Data Mining, Text Mining and Their Business Applications*, vol. 1, 2005, Skiathos, pp. 321–329.

18. Liang, L., Zha, Y., (2014), "Aggregated ratio analysis in DEA", *International Journal of Information Technology and Decision Making*, available online.

19. Marroquin, M., M. Pena, C. Castro, J. Castro, M. Cabrera-Rios, (2008), "Use of data envelopment analysis and clustering in multiple criteria optimization", *Intelligent Data Analysis* 12, 89–101.

20. Meimand, M., R.Y. Cavana, R. Laking, (2002), "Using DEA and survival analysis for measuring performance of branches in New Zealand's Accident Compensation Corporation", *Journal of the Operational Research Society*, 53 (3), 303–313.

21. Samoilenko, S., K.M. Osei-Bryson, (2010), "Determining sources of relative inefficiency in heterogeneous samples: Methodology using Cluster Analysis, DEA and Neural Networks", *European Journal of Operational Research*, 206, 479-487.

22. Schreyögg, J., C. von Reitzenstein, (2008), "Strategic groups and performance differences among academic medical centers", *Health Care Management Review*, 33 (3), 225–233.

23. Sharma, M.J., S.J. Yu, (2009), "Performance based stratification and clustering for benchmarking of container terminals", *Expert Systems with Applications*, 36 (3), 5016–5022.

24. Shin, H.W., S.Y. Sohn, (2004), "Multi-attribute scoring method for mobile telecommunication subscribers", *Expert Systems with Applications*, 26 (3), 363–368.