NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Irina Abankina, Fuad Aleskerov, Veronika Belousova, Leonid Gokhberg, Kirill Zinkovsky, Sofya Kiselgof, Vsevolod Petrushchenko, Sergey Shvydun

PERFORMANCE-BASED TYPOLOGY OF UNIVERSITIES: EVIDENCE FROM RUSSIA

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: SCIENCE, TECHNOLOGY AND INNOVATION

WP BRP 33/STI/2015

Irina Abankina[1], Fuad Aleskerov[2], Veronika Belousova[3], Leonid Gokhberg[4], Kirill Zinkovsky[5], Sofya Kiselgof[6], Vsevolod Petrushchenko[7], Sergey Shvydun[8]

PERFORMANCE-BASED TYPOLOGY OF UNIVERSITIES: EVIDENCE FROM RUSSIA[9]

In recent decades, increased economic pressure and growing societal expectations have led to the introduction of performance-based funding models of public research, namely universities. In this respect, universities have started to use various strategies to adapt and develop their activities under the new framework. National governments are currently attempting to design and apply various taxonomies for structuring university infrastructure in different ways in order to facilitate the development of efficient programmes for the advancement of higher education.

This paper provides a review of different approaches to university typologies, discusses the choice of indicators and mathematical tools for grouping universities using common criteria and evaluating their performance based on classical and modified DEA approaches. The authors develop a typology which was tested in the Russian context, taking into account indicators of research and educational activities implemented by domestic universities and their efficiency score. The typology is based on clustering universities by the availability of resources and research and educational performance and the combination of these results with their efficiency score. It groups universities by type and includes a decision tree for classifying them taking into account their heterogeneity. It serves as a basis for the content analysis of a wide range of universities, and for shaping targeted policies aimed at particular groups.

***Keywords*:** higher education institutions (HEIs), typology, research and educational activities of HEIs, hierarchical clustering, data envelopment analysis, efficiency, performance.

JEL code: C14, C38, D83, O32

---

[1]Director, Institute for Educational Studies, National Research University Higher School of Economics, E-mail: abankinai@hse.ru

[2]Head, Department of Mathematics, Faculty of Economics, and Head, International Laboratory of Decision Choice and Analysis, National Research University Higher School of Economics, E-mail: alesk@hse.ru

[3]Head, Department for Methodology of Budget Planning, Institute for Statistical Studies and Economics of Knowledge, National Research University Higher School of Economics, E-mail: vbelousova@hse.ru

[4]Director, Institute for Statistical Studies and Economics of Knowledge, First Vice-Rector, National Research University Higher School of Economics, E-mail: lgokhberg@hse.ru

[5]Deputy Director, Institute for Educational Studies, National Research University Higher School of Economics, E-mail: kzinkovsky@hse.ru

[6]Lecturer, Department of Mathematics, Faculty of Economics. E-mail: skiselgof@hse.ru

[7]Research Assistant, International Laboratory of Decision Choice and Analysis, National Research University Higher School of Economics, E-mail: vvpetruschenko@edu.hse.ru

[8]Research Assistant, International Laboratory of Decision Choice and Analysis, National Research University Higher School of Economics, E-mail: shvydun@hse.ru

[9] The financial support from the Government of the Russian Federation within the framework of the Basic Research Program in 2014 at the National Research University Higher School of Economics and within the framework of implementation of the 5-100 Programme Roadmap of the National Research University Higher School of Economics is acknowledged.

Introduction

Over the last decade the Russian education system has transformed significantly under the influence of changes occurring at national, regional and international levels including the destruction of the totalitarian system and the disintegration of the Soviet Union, the ongoing globalization of the national science and education systems, the Bologna process and the creation of a common European educational space. The effect of these processes is continuous and to a greater or lesser extent is reflected in the education system transformations. Of all sectors of the Russian education system, higher education is experiencing the most complex and profound changes, and has attracted the most attention from both the government and the general public. Changes in higher education in Russia can be called radical, but they are yet not complete.

The education system in the Soviet Union was centralized and uniform, consisting exclusively of public institutions financed by the state. The higher education system was shaped in the 1930s, in a period of intensive industrialization. During this period, a significant number of state universities were created to provide large-scale education and training programmes to supply engineers to growing industrial sectors. Newly-established higher education institutions (HEI) were focused mostly on education, vis-à-vis a few major classical universities which inherited the traditions of research excellence from the Tsarist era [*Gokhberg et al.*, 1997]. In turn, mass training of pedagogical faculty provided a transition to compulsory secondary education in the Soviet Union, thus colleges of education and teacher training institutes were established in most regions of the country.

In the same period, a system of technical universities was formed for agriculture, transport and communications, health care, etc. These universities were a reaction to the challenge of industrial development. In the late 1980s, enrolment in higher education in the Soviet Union was about 20% of school graduates, while others had vocational training or entered the labour market immediately after secondary school. Despite the fairly elaborate higher education system, the proportion of people with university diplomas did not exceed 10-12% of the total workforce in the Soviet period [*Education in Figures: Pocket Data Book*, 2013, p. 11].

The situation dramatically changed in the post-Soviet era. Economic transformations due to the transition to a market economy and the emergence of private businesses, the growth of new industries such as finance, IT, real estate, professional services, advertising and media, all demanded well educated human resources. It became obvious that university-degree holders were more adaptable and capable of retraining and getting a new job. In the 1990s, Russia experienced a boom of fee paying education services, both private and public. At the end of 2011, the HEI network already involved 1080 universities, 634 of them state-owned and 446 private, when HEIs enrolled 6.5 million students [*Abankina & Scherbakova*, 2013].

Historically with minor exceptions, Russian R&D activities were concentrated outside universities, mostly at the R&D institutions governed by the Russian Academy of Sciences and major industrial agencies[10]. Recently, the government has made significant efforts to foster R&D in universities and increasing their global competitiveness. According to legislation, two new types of universities were specified in Russia: (1) Federal Universities, founded by merging several universities in major regional cities (there are already 9 and more are planned), (2) National Research Universities selected on a competitive basis and focused on cutting-edge research areas. Both groups receive funding earmarked from the state budget for their individual development programs.

The organisational landscape of higher education has significantly changed in recent decades, becoming much more diverse. This is reflected in the increasingly complex organisational structures of universities, the increased number and diversity of educational programmes, HEI expansion into international markets, and the strengthening of R&D, public service, entrepreneurial initiatives. However, these recent changes in the institutional landscape are not reflected in the respective state policies and broadly speaking in the societal understanding of the mission of universities. The main purpose of this paper is to present an approach to, and the results of, a university typology.

There is no space for a detailed discussion of mission here but we provide a brief overview. One point of view, following Humboldt (1903) and Flexner (1930), is that universities should be considered as places of research and measured by their contribution to science. Another view, following Newman (1852) and Gasset (1944) and many others, argues that the main mission of a university is education. The third mission linked to public service is considered as important in a diverse democratic society and equally important to the other missions of university [*Checkoway,* 2001]. Finally, there is the approach which considers every university a unique organization combining many missions [*Boyer,* 1990; *Marginson,* 2007].

The evidence of the gap between public, governmental and scholarly views on universities can be seen in many debates around the existing national and international rankings. The main and most popular international rankings mostly stress the research mission of universities, which is why the government and public cannot distinguish between the various types of universities and their missions. Therefore, there is a need for developing and promoting relevant typologies to support the understanding of university diversity and provide stakeholders with appropriate tools to compare them.

[10] The share of HEIs in R&D expenditure in Russia had been around 5-6% of the national total since the Soviet times until early 2010s [*Gokhberg et al.,* 2011].

First, we make a comparative study of the typologies widely used in the European and American educational landscapes. Then, we develop a typology of Russian public universities using indicators of research and education. Finally, we present the typology of universities combined with their performance score and taking into account the heterogeneity of universities.

The paper is structured as follows. Section 1 provides a comparative study of approaches for university classification and performance evaluation. Section 2 describes the dataset, the classification criteria, and the performance indicators selected. Section 3 presents the methodology for clustering universities and their performance evaluation. Section 4 discusses the proposed typology of universities based on the clustering results. Section 5 shows how the performance of universities is measured and compares the distribution of HEIs by cluster and efficiency score. The final section concludes.

1. Literature Overview

As the paper aims at developing a typology of Russian universities by incorporating their performance score, two separate literature overviews are needed. The first derives a classification of HEIs by showing statistically their similarities and differences. The second takes into account the heterogeneity of HEIs and ranks them on a performance score illustrating an efficient mix of inputs and outputs.

1.1 Approaches for developing a typology of HEIs

International practices demonstrate two radically different approaches to classifying HEIs on the basis of their current performance or potential for strategic development. In the first case, universities are grouped according to their activity parameters, the level of education they provide, the range of disciplines [*Abankina et al.*, 2013b]. For example, the UK traditionally uses various university classifications based on their age and history: Oxford University, Cambridge University, the University of London, technology universities and former technology colleges, new universities (former polytechnics), universities established during the Victorian era ("redbricks"), or new universities founded in the 1960s [*Scott*, 1995]. Another example is the well-known university typology based on research developed by the Carnegie Foundation in the USA in 1970s [*McCormick & Zhao*, 2005].

In the latter approach, universities are grouped by taking into account opportunities and aspirations to grow [*Abankina et al.*, 2013b]. For instance, unlike the traditional classification of British universities, the University of Manchester developed one unrelated to HEIs history and other *a priori* characteristics such as size, research, third mission, social activities and availability, and training [*Howells et al.*, 2008]. The latest version (2005) of the Carnegie typology is other example which includes a number of parameters each of which classifies a university into a certain group, taking into account multiple aspects of their activities. Currently, the classification uses criteria such as the quality of education and research, type of teaching differentiated separately for Bachelor and Master/PhD programmes, student profile, size and environment [*McCormick & Zhao*, 2005].

The latter approach to HEI classification has become more popular, evaluating only the current performance, incorporating lifelong learning educational programmes, R&D, and innovative projects. Moreover, universities which strive to grow but have not yet reached the required level, are not branded as low-quality, which positively affects their development prospects [*Abankina et al.*, 2013b]. For example, the University of Twente implemented a study sponsored by the European Commission aimed at developing a pan-European typology alongside the results of the Carnegie classification [*University of Twente*, 2005]. A wide set of parameters on education (teaching and training, student enrolment), R&D, regional networking, participation in knowledge exchanges and international activities served as a basis for harmonised pan-European classification [*Bartelse & van Vught*, 2007]. The typology is country-specific and includes data collection via special monitoring programmes, based on voluntary participation in a survey. A threshold condition for university inclusion is having at least one government-certified educational programme. This classifies 67 European Institutions for Higher Education for the European Commission and maps their diversity [*CHEPS*, 2008].

Furthermore, special attention has been paid to the following dimensions to characterize the European landscape of universities:

- growth dynamics,
- specialization pattern,
- subject mix,
- funding composition,
- proportion of post-graduate students and
- productivity [*Daraio et al.*, 2011].

In spite of the broadness of the indicators, the authors found uncorrelated outputs of the universities such as research, funding and top researchers based on a sample of 488 European HEIs in 11 countries. Nevertheless, they argued that the differentiation of European universities occurred in some cases, in particular with respect to research output and competitive funding.

At the same time, there are two radically different methods for constructing a typology: either based on expert selection of the classification criteria and threshold parameter values for determining university types, or on mathematical clustering models [*Abankina et al.*, 2013b].

The latter groups universities in a way that they turn out to be sufficiently uniform, and significantly different from each other[11]. As a rule, universities are grouped by the parameters in question via clustering. There are some examples in the EU, UK, Philippines and Russia [*Bonaccorsi & Daraio*, 2009; *Howells et al.*, 2008; *Bernardo*, 2003; *Abankina et al.*, 2013b].

In turn, Bonaccorsi & Daraio (2009) identified the dominant cluster (77% of the European sample) as between the research and the teaching objectives of universities, i.e. an intermediate group. Only 12% out of 271 observations were classified as research-orientated HEIs. In addition, the authors found that the size of HEIs was the most significant indicator for clustering. However, they also showed that HEIs in the UK, the Netherlands and Switzerland differed not just based on size-related indicators; they were characterized by variation of research and teaching intensity. They also pointed out that country effects still mattered for the research intensity heterogeneity of HEIs.

Similarly, Bernardo (2003) primarily focused on the size of 220 HEIs in The Philippines and split them into three aggregate groups: large, graduate-capable (master), and small universities. At the same time, three universities were considered to be unique and not grouped with any of the others. For each university in the Philippines, data for 45 indicators were collected, describing various aspects of their activities. These included, among others:

- the year of establishment,
- the number of educational programmes and their shares in the total number, by level (Bachelor’s, Master’s, Post-graduate Studies, upgrading and retraining),
- the number of students enrolled in programmes, and their shares in the total cohort, by the above levels,
- the number of graduates and their shares in the total cohort, by the above levels, cohort of faculty members—degree holders, by level (PhD, Master, Bachelor), their total number,
- the number of publications by faculty members, in international and selected Philippino academic journals (in total, and per faculty member).

12 clusters were formed, each containing 4 subgroups. The first subgroup of Large universities included two HEIs, with no great variety of educational programmes. Compared with other subgroups, it showed a large share of Master’s students in the total number of students, and low publication activity by faculty members. The second subgroup had 3 HEIs and offered a wide range of educational programmes, mostly at Bachelor and post-graduate levels.

[11]The preliminary findings for clustering can be found even when HEIs are compared just by constructing the national boxplots, i.e. analyzing range, median and quartiles of the indicators within and across countries [*Daraio et al*, 2011].

Faculty publication activity was higher than in the previous subgroup, and most faculty members had a degree. The third subgroup of 5 HEIs had practically no Master's programmes on offer and no more than 3% of teachers with degrees. However, they had the largest number of students. Finally, the last subgroup, with 9 HEIs, was very similar to the previous subgroup but provided a wider choice of educational programmes and a higher share of faculty had degrees.

Compared with other universities in the category of Master universities, these 17 HEIs in the first subgroup were relatively small, offered a choice of Master's programmes, and conducted a limited amount of R&D. In the second subgroup, the 15 HEIs were much bigger, the share of masters and post-graduates in the total number of students was significant, while the level of R&D remained extremely low. The third subgroup had 13 average sized HEIs (in between subgroups 1 and 2). The share of master's programmes was slightly smaller than in previous groups, while the share of post-graduate programmes was about the same. The fourth group consisted of 30 quite small universities with a lower share of Master's and post-graduate programmes than other Master universities. The share of teachers with degrees was also rather small. At the same time, this group demonstrated the best R&D results among all Master universities.

The Small universities category included four subgroups of institutions with small numbers of both students and teachers, offering a limited choice of educational programmes. For example, the first subgroup of 24 HEIs offered a minimal choice of educational programmes, mostly Bachelor-level, but boasted relatively high-quality teachers (about 40% of faculty members had a degree). In contrast, the second subgroup of 11 HEIs had a high share of Master's programmes and a relatively low level of teachers (no more than 20% of them have degrees). The third subgroup of 37 HEIs was similar to the previous one, but had slightly higher values for the total number of faculty and the proportion of degree holders. The last subgroup had 54 HEIs with the smallest size and an appropriately small choice of programmes, exclusively at bachelor level. The share of teachers with degrees was on average about 30%.

Bernardo (2003) examined the orientation of education programmes offered by the HEIs as the differences between the above groups, which are mostly based on the educational programmes and education level of teachers. The clustering in question reveals just one layer of the Philippino higher education system because it does not cover universities catering for different target audiences. Breaking universities down by the quality of their research is largely dichotomic, because of the use of quite stringent evaluation criteria of publications in peer-reviewed journals.

In contrast to Bonaccorsi & Daraio (2009) and Bernardo (2003), Howells et al. (2008) selected the number of clusters in the 2-20 clusters range using the Caliński-Harabasz index

[*Caliński & Harabasz*, 1974]. This is calculated as the ratio of the overall spread of objects between clusters, and inside them. The number of clusters which maximises the index of the value is considered to be optimal. In the end, seven clusters were identified: universities oriented towards applied research, locally available universities, research elite, “professionals for London”, active recruiters of students, universities oriented towards R&D and actively recruiting students, Open University as a separate cluster.

The key variables in the typology of Howells et al. (2008) describe the multidimensional change of universities in size, research, teaching, third mission and social inclusion. Such a division into uniform clusters may help in the design of government higher education policy, e.g., it allows the government to identify universities primarily oriented towards education, and discover where there are high concentrations of students in need of social support. From the perspective of research and innovative activities, universities which should be first to receive government support are moved to the foreground.

The Russian evidence suggests that universities should be classified not just by formal characteristics, but also by other indicators which shed light on their mission and strategic development. Empirical typologies were developed based on adaptation strategies (defined on the basis of resource availability and output, see Appendix 1) [*Titova*, 2008], economic models [*Abankina et al.*, 2010] and a combination of their activities (R&D, education) [*Abankina et al.*, 2013b].

In the Russian context the most interesting typologies to researchers and practitioners reflect R&D and education activities. These aspects can be covered by an input and output mix (product characteristics and availability of resources). Such typologies create a multidimensional picture of various strategies and their efficiency, hence the potential may be measured by the means of data envelopment analysis (DEA).

1.2 Data Envelopment Analysis as a tool for performance measure of heterogeneous HEIs

We start with some issues which have been identified with DEA before applying it to the Russian dataset. Firstly, the context of input and output specification has attracted a great deal of scholarly attention. For example, Johnes (2006) had some concerns with the definition and measurement of input and output bundles for HEIs, and their importance in the DEA specification. Secondly, researchers should be sure which inputs are under the control of HEIs. To account for this, Kempkesa and Pohl (2010) and McMillan and Wing (2006) used DEA with environmental variables. The difficulties with a set of inputs and outputs arise from the peculiarities of the higher education production process.

To the best of our knowledge, DEA is less used in the Russian higher education context. Abankina et al. (2012) did a pilot project on 30 Russian HEIs using a classical DEA model with constant return to scale. The authors found that the model for a university funding structure (institutional vs. competitive based ones) provided more comparable results to the international evidence. These results were confirmed in the subsamples of 78 technical and 59 classical universities in [*Abankina et al.*, 2013a].

However, when calculating efficiency scores of HEIs, there is a challenge related to the heterogeneity of the sample. This may lead to an inaccurate estimation of HEI efficiency due to the identification of incomparable peer groups. To overcome this problem, many authors used DEA separately for each homogenous subsample. For example, Beasley (1995) distinguished among units by the goals separately set up for and achieved by research and education activities. In turn, Abbott and Doucouliagos (2003) represented the ratio of grants received to the number of students in full-time equivalent to divide the sample into three clusters and assess the efficiency of each of them. Warning (2004) distinguished between the research and education strategic goals of the units and their specialization. Johnes and Yu (2008) found that the funding modes and locations of Chinese universities played a crucial role in explaining the differences among their efficiency values. Finally, Athanassopoulos and Shale (1997), similarly to Johnes (2006), introduced a specialization criteria to have three similar subgroups of British HEIs.

In addition to specifying criteria for sample heterogeneity, some authors propose modifying classical DEA. A recent study [*Johnes et al.*, 2011] applied DEA to the cost structure, efficiency and productivity of British English universities. The authors used a division into clusters similar to Johnes (2006) and then run DEA within each cluster separately. By contrast, Athanassopoulos and Shale (1997) used DEA stage-by-stage. In this algorithm [*Coelli et al.*, 2005], the first subgroup was evaluated, then the first and second ones were together, and finally, the full sample was assessed. The efficiency scores of the *i*-th cluster were obtained at the *i*-th algorithm stage. Aleskerov and Petrushchenko (2013) developed another algorithm and applied it to 29 Russian universities. This was based on a sequential exclusion of alternatives. The procedure reduced the influence of outliers on the efficiency evaluation.

2. Data description and indicator selection

Our typology was constructed on the basis of the data collected in the course of standardized written surveys of university R&D and innovation activities and their cooperation with companies, conducted by HSE’s Institute for Statistical Studies and Economics of Knowledge, in the framework of the Economics of Science Monitoring Project covering 400 state universities in Russia. These data were supplemented with the results of Unified State Exam (USE) across universities found in open sources[12].

However, not all the HEIs provided data on the indicators used and outliers were excluded from the sample. As a result, 219 HEIs were analysed. Despite the reduced list, it still featured more than half of Russian universities from the sample, ensuring the reliability of the results.

The characteristics for clustering were selected based on the objectives of constructing a typology to identify groups of universities with a similar potential for innovation development, research and education activities. The choice of indicators is supported by the theoretical foundations of Bernardo (2003), Bartelse and van Vught (2007), Howells et al. (2008), Bonaccorsi & Daraio (2009), Abankina et al. (2013b). This generates the benefit of clustering Russian HEIs by a combination of their activities (research, education, innovation). Notably, growing attention has been given to the dependence of HEIs on public funding and the diversification of revenues, which indirectly describe university types via R&D funding from non-state sources, i.e. this shows how active and prosperous the university is in R&D activities.

At the same time, specific attention was paid to the peculiar features of the available data. First, excluding parameters with a significant statistical connection or correlation between them. Second, some indicators were disregarded due to numerous data errors and outliers. E.g. the number of university faculty was frequently underreported, so all characteristics based on this value were excluded. The resulting clusters differ by at least one indicator.

According to our calculations, the share of R&D expenditure in the total revenue correlated with the absolute value of non-state funding provided for these activities at 0.65 (significantly different from zero), so the second parameter was disregarded. The total number of students and the number of students per teacher also have positive correlation (approximately 0.5). The second indicator was excluded due to obviously faulty data on faculty numbers provided by certain universities. Finally, the number of scientific papers per faculty member, and the cohort of students per teacher have positive correlation of 0.68. Presumably, the smaller the total number of teachers, the more papers each publish, and the weaker the effect of "publication diffusion". To calculate a specific publication activity indicator, we used the total number of teachers and researchers on the university's payroll (including freelancers) as denominator.

Finally, the following five parameters were chosen for clustering:

1. USE – average USE grades of the students enrolled in the relevant year;

[12] This is an obligatory exam for every student to both graduate from schools and enter any university or college. Since 2009, it has been introduced instead of preliminary entrance examinations at universities. For more details please refer to "Public monitoring of university enrollment procedures as a condition of providing equal access to education". Access mode: http://www.vsu.ru/news/index.do?id=3164 (last accessed on 16.10.2014).

2. Youngsters – share of young faculty members (for those under 30 without a degree, under 35 for Candidates of Science (PhDs), and under 40 for Doctors of Science) in the total number of faculty;
3. Unit Public Funding (UPF) – public funding per student provided in the relevant year;
4. Size – total number of students (publicly funded and privately funded places);
5. R&D – share of R&D funding obtained from non-state sources in the total expenditures for R&D.

Despite the small number of parameters, clustering reveals university potential in major areas. UPF describes education and research activities incorporating all public spending, including science and education expenditures. The USE parameter reflects university educational potential and the potential involvement of students in R&D. The share of Candidate of Science (like PhD) and Doctor of Science (like German Habilitation) degree holders in the total number of faculty members reflects the potential for human capital reproduction, and related development prospects. The share of non-state funding in R&D expenditure measures and investments in applied research reflects the overall financial situation of the university. Finally, the university size is evidence of its economic stability (as the public funding of Russian universities is proportional to the number of students) but at the same time it can be evidence of lower education quality and lower student involvement in R&D. The presence of talented students, a balanced faculty structure and solid financial resources should indicate a good R&D performance, and provide additional opportunities to upgrade R&D activities [*Abankina et al.*, 2013b].

Bonaccorsi & Daraio (2009) suggested incorporating data on discipline differentiation when providing a full-scale classification of HEIs. While Abankina et al. (2013a) tested this only for subsamples of technical and classical universities, we calculate the efficiency score for all subsamples available, i.e. 53 classical, 29 pedagogical, 24 socio-economic, 70 technical, 10 architectural, and 11 humanities HEIs[13]. Then, we conduct a comparative study of their educational (Model 1) and academic (Model 2) potential. The specifications of both models are in line with Abankina et al. (2013a).

In Model 1, the educational potential of Russian HEIs is tested. The following input parameters are used: UPF, share of teachers with degrees (PDS), total number of teachers (PPS). Output parameters included first year student average USE mark (USE) and total number of students (STUD). This set of output indicators characterizes student choice of a particular university given its quality of education.

[13] We have less than 219 HEIs in the sample for DEA analysis due to outliers within these groups: 6 classical, 4 pedagogical, 1 socio-economic, 8 technical, and 2 architectural HEIs, and 1 the humanities.

Model 2 measures the academic potential of HEIs. The input parameters are similar to Model 1 with the addition of the first year students' average USE. The last parameter is used as an indicator of the academic potential of entrants, who can participate in university research projects. Output parameters include the total number of students (as in Model 1) and the share of young teachers (YPPS) and publication activity of the faculty members (PUB).

The descriptive statistics of the dataset for clustering and performance evaluating are presented in Appendix 2 and 3, respectively.

3. Methodological Choices

3.1 Methodology of clustering HEIs

Calculations were performed using a hierarchical clustering technique which constructs a tree based on the initial set of objects (for more details see Appendix 4). The advantages of this method compared with k-means clustering (used, for example, in [*Howells et al.,* 2008; *Bonaccorsi & Daraio,* 2009; *Abankina et al.,* 2013b]) are that it provides extra information about the structure of the data set and allows an endogenous choice of the number of clusters based on the characteristics of the clusters. Moreover, breaking objects down into uniform groups using a k-means clustering technique depends on the initial selection of the clusters' centre position. Since objects are grouped around the centres, we can get variations depending on the *a priori* assumptions about the initial positions of cluster centres.

We have considered clustering techniques oriented towards analysing quantitative characteristics of objects. There are also numerous methods designed to work with dichotomic variables. Whether it is possible or appropriate to convert university characteristics into a dichotomic format to perform clustering on that basis is an issue for future research. In our case, each variable was initially converted into the 0–100 range. Such an approach to normalisation is widely used and is one of the most preferred [*Mirkin,* 2011] as it does not introduce structural changes as, for example, a standardizing technique does.

All variants of hierarchical approaches are based on various techniques for calculating the distance between two isolated objects, or between an object and a cluster. These determine clustering results, so the choice of calculation technique requires particular care. We used a Euclidean distance in five-dimensional space to measure the differences between objects.

In our case, the Euclidean distance proved to be the most reliable and appropriate metric. It allows us to ignore small differences between objects, and at the same time imposes heavy fines for significant differences for even a single parameter. We assume that universities which belong to the same cluster must be as uniform as possible for all the selected criteria. Accordingly, we chose a Euclidean distance instead of a metric which would take into account only marginal (maximum and minimum) divergence between objects, by all characteristics. This alternative metric is also turned very sensitive to outliers.

The key cluster characteristic (preliminarily in the tree, and finally at the analysis of the object distribution stage) is the internal spread of values. In this case the spread is measured via the sum of the distances from the objects to the cluster's centre. The nature of the data was taken into account when the spread measure was selected. Distance from centroid to the farthest object (which in another case could ensure a more uniform distribution), in our case turned out to be an unreliable measure due to its high sensitivity to outliers.

Accordingly, we used the Ward method [*Ward*, 1963] to measure differences between clusters and objects, or between two clusters. In this case, the distance between two clusters is seen as the increment of the total distances between all grouped objects and the centre of the cluster, compared with total distances between objects to the centre in each individual cluster. In other words, we measure the deterioration of the internal cluster spread which happens if two clusters are merged.

A pair of clusters with minimum divergence, in the above sense, is selected for merging as a specific feature of the Ward method is a tendency to produce clusters of comparable size, because smaller clusters join larger ones. In our study, this feature allowed us to obtain more reliable and sustainable results, as it allowed us to identify different university types instead of producing a heterogeneous distribution with small clusters of unique objects.

The Ward method produces a clustering with low variance of objects inside the clusters. The product is a hierarchy of object groups, called a taxonomy. Therefore, this technique not only breaks the sample down into clusters, but also shows the data structure. After the taxonomy is built, the tree is cut off at a certain distance. All distances between clusters, which are smaller than a selected threshold value, are considered to be merged. We have analysed the clustering of the universities after cutting the tree off in various ways. We selected a cut-off resulting in 6 clusters, described in more detail in the next section.

3.2 Methodology of estimating HEIs performance

To evaluate the efficiency score of HEIs, two techniques are applied: econometric (stochastic frontier approach (SFA))[14] and linear programming (data envelopment analysis (DEA)) methods. Under DEA, the functional frontier parameters are not estimated, their significance is not tested, and the distribution of inefficiencies is not assumed. Instead, linear programming methods are used to compute a piece-wise surface over the available observations.

[14] The paper does not cover the statistical one. The basics of this approach are given in the book [*Coelli et al*, 2005].

DEA calculates an efficiency score for each HEI in comparison with the best ones in the dataset relative to this given frontier. This score is introduced via the ratio of the set of outputs over the mix of inputs. Optimal weights for outputs and inputs are found by solving the linear programming problem, which is visualized by introducing a simple example under the assumption of a constant return to scale in Appendix 5. To introduce an equivalent envelopment form, dual linear programming is employed [for more details see *Charnes, Cooper & Rhodes*, 1978]:

$$\min_{\theta_k,\lambda} \theta_k$$

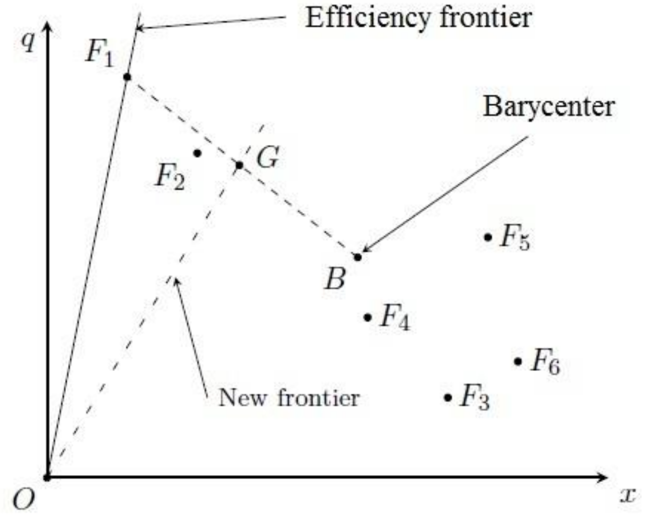
subject to

$$\begin{cases} -q_k + Q \cdot \lambda \geq 0; \\ \theta_k x_k - X \cdot \lambda \geq 0; \\ \lambda \geq 0, \end{cases} \qquad (1)$$

where q_k is a vector of all output weights for the *k*-th HEI, x_k is a similar vector of input weights, *Q* is a matrix for outputs, which columns are $q_1, \dots, q_k, \dots, q_L$, where *L* is the sample size, *X* is a matrix for inputs, λ is a vector of constants and θ_k is a value from [0,1]. This scalar represents an efficiency score of the *k*-th HEI. When it equals 1, it identifies an efficient HEI lying on the piece-wise surface. This problem must be solved *L* times for each unit in the sample.

In turn, the model proposed by Aleskerov and Petrushchenko (2013) is based on the standard DEA technique. The method can be illustrated by one input-output mix (Fig. 1). Given a set $F_1, \dots, F_6$ of HEIs, the barycentre *B* is calculated. The frontier is moved toward point *G* defined as μB + $(1-\mu)F_1$, where $\mu \in [0,1]$ is the parameter of heterogeneity. Then, using DEA we can estimate the efficiency score of all appropriate HEIs ($F_3, \dots, F_6$) via the new frontier generated by *G*. The process continues until all units are evaluated. Generalization of the algorithm to the case of an arbitrary number of inputs and outputs is given in [*Aleskerov & Petrushchenko*, 2013].

Figure 1. DEA by sequential exclusion of alternatives.



4. Clustering results of Russian HEIs

Table 1 presents the quantitative parameters of each cluster centre, and the number of objects in them. Clusters with average indicator values significantly above the overall average are highlighted in dark blue. Clusters with the appropriate average indicator values significantly below the overall average are highlighted in light blue.

Table 1. Cluster profiles and object distribution

Cluster number	USE	Youngsters	UPF	Size	R&D	Number of objects
1	59.81	0.24	82.75	5905.31	89.82	36
2	66.32	0.15	90.05	13495.43	86.92	37
3	57.42	0.06	95.82	5825.06	71.61	32
4	70.77	0.16	205.40	7947.39	70.72	18
5	59.53	0.21	85.15	7177.98	50.80	42
6	65.00	0.15	101.54	9005.85	13.41	54

Source: [*Abankina et al.*, 2013b]

Table 1 shows that by the average value of first-year student USE marks, clusters can be divided into two groups. Universities in clusters 1, 3, and 5 show statistically significant lower results than universities in clusters 2, 4, and 6. Deviation between clusters within each group is insignificant.

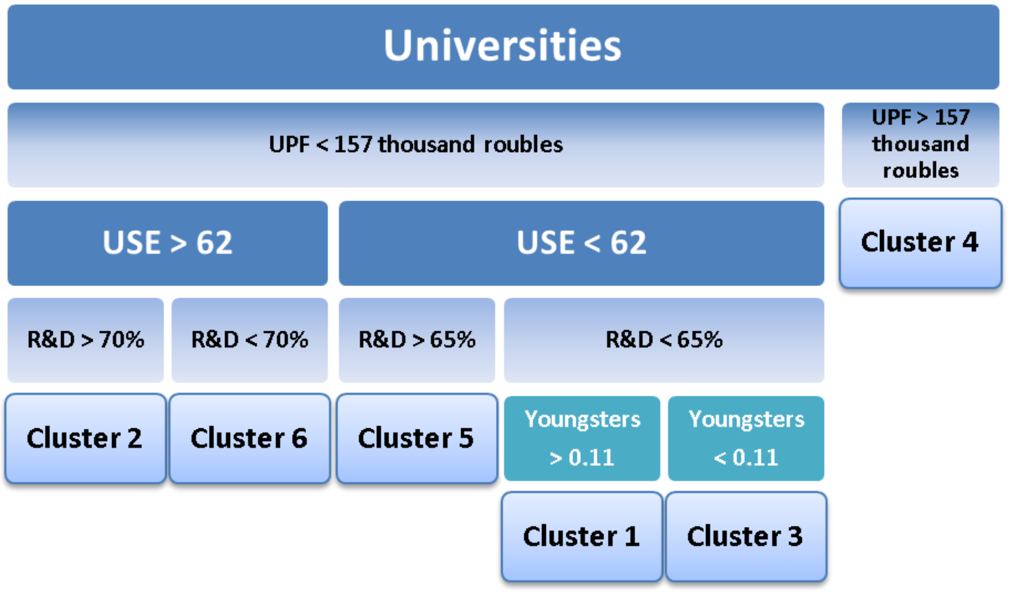
By the share of young faculty members in the total number of personnel, the clusters (taking into account deviation significance) range in the following way: universities in cluster 3 have the smallest values averaging about 6%. Then follow clusters 2, 4, and 6, with approximately 15%, and cluster 5 (21%). Cluster 1 has the highest share of young faculty members (24%). In terms of unit public funding, cluster 4 stands apart from all the others, which are relatively uniform.

The largest universities were in clusters 2 and 6 (significantly bigger than all others); medium-size universities were in clusters 4 and 5, while the smallest (with about 5,000 students) were in clusters 1 and 3.

In terms of the share of non-government funding in R&D expenditures, universities in clusters 1 and 2 stand out. Practically all their R&D is funded from non-government sources. In cluster 6 on average only 13% of R&D expenditures are non-government funded. The remaining three clusters (3, 4, and 5) hold intermediate specific positions. Their value of this indicator is much lower than in clusters 1 and 2, and significantly higher than in cluster 6.

Each cluster has a set of indicators whose values make the universities significantly different from all others in the sample. Since clustering does not produce pure types (clusters always overlap in projections on indicator axes), to simplify things we have created a classification tree as a tool for including universities into appropriate groups (see Figure 2 below).

Figure 2. Classification tree



Source: [*Abankina et al.*, 2013b]

These identify five university types grouped into six clusters.

Niche universities (clusters 1 and 3). These are small universities with 3000 to 7500 students, accepting entrants with low USE grades (up to 63%). R&D is mostly financed from non-government sources. For example, for cluster 1, the share of R&D funding obtained from non-state sources in the total R&D expenditures is 90%, for cluster 3 it is 72%. This group has a high share of young teachers and a low UPF value (between 60000 and 90000 roubles). The R&D potential of these universities is limited due to their fixed position in local territorial niches, with limited resources to finance their development and limited prospects of obtaining increased public funding, and a high dependency on their main customers. Half are pedagogical and classical universities, up to one third technical universities.

Indeterminate position universities (cluster 5). In general these are similar to niche ones. They comprise small and medium sized universities (up to 10000 students), with a low quality of entrants, a significant share of young faculty members, insufficient R&D funding (both from public and other sources, in approximately equal shares), and a low UPF value. They are likely to experience reduced enrolment and gradually transform into niche universities. Their R&D potential (in terms of resources) is insignificant.

Market leaders (cluster 2) are large universities which mostly accept students with high USE marks. Their R&D is financed primarily from non-government sources. These universities actively market their educational services. These are primarily classical and socio-economic universities.

Potential and current R&D and education leaders (cluster 4). They receive high government funding per student: more than 157000 roubles; have high average USE scores for new entrants (70%); and boast the highest R&D potential. With their powerful resource base, they can (though not always do) develop and integrate research and educational activities. More than half of these universities specialise in technology.

Universities of good standing (cluster 6) are medium sized or large, and accept high-quality entrants. Public funding in this cluster is on average slightly higher than in other groups, but the cluster is not uniform in this respect. Another feature is the absence of non-government R&D funding. The cluster includes mainly classical and technical universities. Some of them have a long history and are well-known. They enjoy a good reputation among customers and government agencies; good connections and highly skilled faculty. However, this potential may reduce because of their conservatism, which has implications for their ability to serve external customers and work with partners.

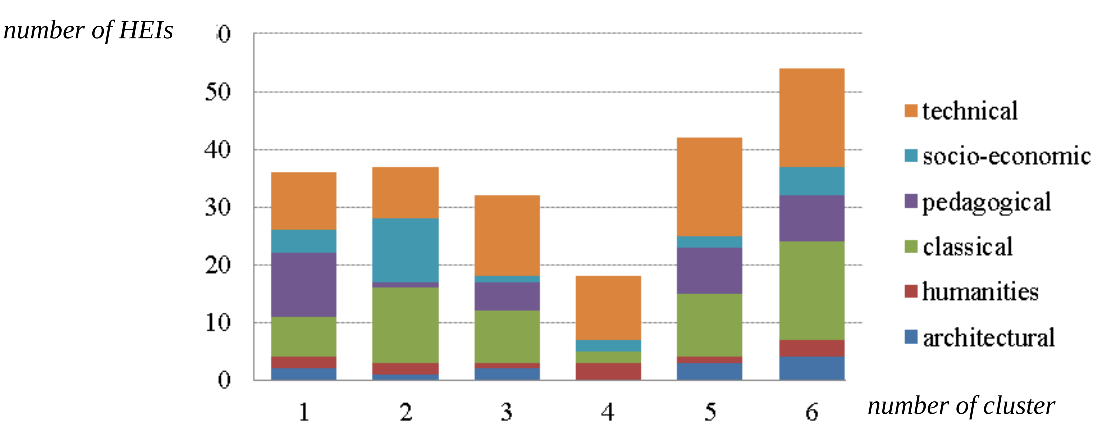
Now let us go back to how the universities under consideration are divided into clusters (Table 1). Almost half of them fell into *niche* and *indeterminate position* clusters, which gives a good idea of the research potential of a large proportion of the country's HEIs[15]. *The potential and current R&D and education leaders* cluster comprises less than 10% of the surveyed universities. They have resources to support R&D activities, but these resources do not guarantee results, and there may be even fewer real leaders among them.

The remaining two groups, *market leaders* and *universities of good standing* comprise about 40% of the universities in the sample. Individual analysis of the universities indicates that they mostly concentrate on education, though they do have certain (in some cases significant) R&D potential. These establishments meet key stakeholder requirements and have good development prospects.

Figure 3 shows the distribution of homogeneous HEI clusters according to the proposed typology. The diagram includes architectural and humanities universities, despite their small number in the sample. This distribution is very much in line with expert expectations, which supports the validity of the proposed typology

- most of the pedagogical universities (about 70%) belong to *niche* and *indeterminate type* clusters;
- a significant proportion of socio-economic universities (about 65%) are represented by *market leaders* and *good standing* clusters; the highest concentration of them is noted in the *market leaders* category;
- almost 50% of technical universities (and about a half of all universities in the sample) belong to *niche* and *indeterminate position* clusters; note that technical universities have a relatively better representation among potential R&D and education leaders (about 15%).

Figure 3. Distribution of various types of universities in the proposed typology



Source: [*Abankina et al., 2013b] and figures for* architectural and humanities HEIs are calculated by the authors.

[15]We would like to stress again that we discuss there sources for universities' R&D activities.

The comparison of the proposed typology with other classifications concludes:

- The availability of resources, and university performance indicators used to design the typologies are similar to the indicators used abroad and which describe various aspects of university activities;
- The proposed classification is similar to the typology of British universities [*Howells et al.,* 2008], in terms of objectives, parameters used, and identified university types.

A drawback of the clustering approach is that each cluster is blurred in the space of the vectors reflecting the availability of resources, which may include universities whose values for this parameter are too different to allow a direct comparison of objects in the same cluster. For correct comparison of universities of the same type, their relative efficiency in the sample should be assessed.

5. Empirical results: distribution of universities by clusters and performance

Distributions of efficiency scores for Models 1 and 2 for the subgroups of universities based on subject mix, are presented graphically in Appendix 6 and the descriptive statistics of the efficiency gains are given in Appendix 7. For each comparable cluster of HEIs with the respect to subject mix, the results obtained from Model 2 seem more convincing than from Model 1. The task is to correctly separate these HEIs with low and average efficiency. HEIs with average efficiency have a development potential based on their own resources, while HEIs with low efficiency scores show poor results in both research and education; they do not attract strong and motivated students and do not have enough financial resources for development.

Moreover, Model 2 allows us to divide all clusters of universities into three subgroups with low, average and high efficiency scores. For example, many well-known technical HEIs are classified as average according to their efficiency scores. It means that these HEIs do not fully use their potential and resources. The subgroup with high efficiency scores include several strong regional HEIs. These HEIs fully use their potential: while having less resources and public funding than some well-known HEIs, they obtain better results.

The comparison of the efficiency scores of Russian and foreign universities shows that Russian universities are on average less efficient than Canadian [*McMillan & Wing,* 2006] and British universities. For example, Johnes (2006) showed that for HEIs specialized in arts and music the average efficiency scores were above 90%. Minimal efficiency score for British HEIs based on their research profile was 66.86% [*Athanassopoulos & Shale,* 1997]. This score is close to the current efficiency score for Russian HEIs (see Appendix 7). According to only educational indicators Russian classical (78.93%), technical (71.43%), pedagogical (90.14%), socio-economic (84.76%), architectural (94.31%), and humanities (84.0%) HEIs are on average more efficient than Chemistry (66.59%) and Physics (64.79%) faculties of British HEIs [*Johnes*, 2006]. However, for both research and educational indicators (Model 2) efficiency scores for Russian HEIs are lower than for Model 1, for instance socio-economic universities for Model 2 have efficiency score no higher than 66%. Russian HEIs are more efficient than German, where there is only 71.9% average efficiency rate in the models specified in [*Warning*, 2004]

For a robustness check, we take into account the heterogeneity of the HEIs as Aleskerov and Petrushchenko (2013) did. Three different values of μ (0.2, 0.5 and 0.8) are introduced. The rankings of universities obtained by this method when μ equals 0.2 or 0.5 are in most cases comparable and statistically consistent with the DEA obtained above (for more details see Appendix 8).

After this, the typology and efficiency of units were analysed, simultaneously. Figure 8 shows the university distribution inside clusters 1-6 by efficiency score calculated by DEA for the two models of HEIs activities (Appendix 9). A large group of universities (clusters 1, 3, and 5) have a limited resource potential for R&D and show low performance figures (using DEA). This is particularly relevant for cluster 5 (*indeterminate position universities*), which confirms the earlier assumption about the shrinking of this type of universities to *niche* size. Here we see university R&D potential and performance limited by default by their resource limitations.

Meanwhile, *market leader* universities (cluster 2), on average show significantly higher performance figures (again using DEA) than other university types. Over 40% of universities in this group can be classified as highly efficient (technical efficiency indicator between 0.8 and 1), and 50% show high performance (very close to 1). This means that *market leaders* get high returns on their resources as *universities of good standing* do. The HEIs from both clusters are able to use their resources more efficiently than universities in other categories, can set sensible objectives, and achieve good results.

At the same time, results for *potential and current R&D and education leaders* are somewhat unexpected. Over 40% of such universities have low performance figures, and under 10% of them can be classified as highly efficient. This is due to the specific DEA model applied.

We tested models with different input and output parameters and find that publication activity used as an output factor significantly affects the measurement of university performance, in particular, technical and classical ones. Publication activity changes the distribution of efficiency measurements: if it is disregarded, the result is a histogram rising with productivity growth and with a certain reduction in the number of universities at the marginal value of performance. However, if it is taken into account, universities are clearly divided into groups with different performance values. Individual analysis of universities reveals that publication activity allows some of them to come close to the efficiency frontier; others are not really affected by this parameter; and others still suffer a significant reduction of their performance figures. Accordingly, when measuring technical efficiency, adjustments should be made for universities who face barriers to open publications if this significantly (and unreasonably) reduces their performance figures. At the same time, for low-performing universities in the group under consideration who do not face publication barriers, the above factor can be explained by their excessive resource potential combined with modest actual results. In other words, the obtained results can help reveal excessive investments in certain universities. This effect will possibly not be observed over a longer period of time.

Conclusions

The analysis of the existing typologies suggests that the most interesting of them are those which reflect various aspects of university activities. These include R&D, educational and innovation activities which can be described not just via results-based indicators, but also through resource availability.

In the latter case, we can identify the university type and its ability to perform certain activities. Such a typology, supplemented with measurements of resource utilisation (e.g. performed with the help of DEA) provides a multidimensional picture of university strategies, and their success rates. Several classifications are based on similar indicators, which gives grounds to define and apply a reduced all-purpose set of characteristics, without a loss of accuracy.

In the most indicative international examples universities were grouped by clustering; the technique also used in our study. With the help of the hierarchical method, we identified and described six clusters. The proposed typology includes a decision tree which divides universities into specific groups with a detailed description.

Processing empirical data about Russian state universities allowed us to break them down into several types, including *niche universities*, *market leaders*, *potential and current R&D and education leaders*, *universities of good standing*, and *indeterminate position universities*.

Due to the historical and modern context of the Russian higher education system, a more diverse organizational landscape might have been expected. This could possibly be explained by the non-dynamic parameters used for clustering. Another reason could be the relatively small set of adaptation strategies used by Russian universities.

The typology allows an analysis of the diverse corpus of Russian universities and the development of targeted policies for each of the identified clusters. Supplemented with performance indicators measured using DEA, the typology provides a better understanding of each cluster's situation, and the formation of strategy for each specific university. The proposed approach to constructing typologies and measure university R&D potentials, generates nontrivial results and provides for the shaping of targeted higher education policies.

The typology presented here takes into account various aspects of university activities, and is not an analogy of the existing ratings or other performance indicators described in the literature. Improving the system for collecting data about university performance and refining the quality of data would facilitate further research in this field.

References

Abankina I., Abankina T., Nikolaenko E. (2010) Ekonomicheskoe polozhenie vuzov v usloviyakh byudzhetnoi reformy, povysheniya avtonomii i vvedeniya EGE [The economic situation in universities in terms of budget reform, improving the autonomy and the introduction of the USE], Moscow: HSE.

Abankina I., Aleskerov F., Belousova V., Gokhberg L., Zinkovsky K., Kiselgof S., Shvydun S. (2013b) Tipologiya i analiz nauchno-obrazovatel'noy rezul'tativnosti rossiyskikh vuzov [A Typology and Analysis of Russian Universities' Performance in Education and Research]. Foresight-Russia, vol. 7, no 3, pp. 48-63 (in Russian)

Abankina I., Aleskerov F., Belousova V., Zinkovsky K., Petrushchenko S. (2013a) Otsenka rezul'tativnosti universitetov s pomoshch'yu obolochechnogo analiza dannykh [Evaluating Performance of Universities Using Data Envelopment Analysis]. Voprosy obrazovaniya, no2, pp. 15-48 (in Russian)

Abankina I., Aleskerov F., Belousova V., Bonch-Osmolovskaya A., Petruschenko V., Ogorodniychuk D., Yakuba V., Zin'kovsky K. (2012) University efficiency evaluation with using its reputational component / Proceedings of the 4th International Conference on Applied Operational Research. Bangkok: Tadbir Operational Research Group. P. 244–253

Abankina I. V., Scherbakova I. (2013) Russian Higher Education Reforms and the Bologna Process, Journal of the European Higher Education Area. No. 3, pp. 3 -25.

Abbott M., Doucouliagos C. (2003) The efficiency of Australian universities: a data envelopment analysis, Economics of Education Review, №22, pp. 89 – 97.

Aleskerov F., Petrushchenko V. (2013) DEA by sequential exclusion of alternatives. WP7/2013/02. Moscow: Publishing House of the Higher School of Economics. – 28 p.

Athanassopoulos A. D., Shale E. (1997) Assessing the Comparative Efficiency of Higher Education Institutions in the UK by the Means of Data Envelopment Analysis, Education Economics, №5, pp. 117 – 134.

Bartelse J., van Vught F. (2007) Institutional Profiles: Towards a Typology of Higher Education Institutions. IAU Horizons, vol. 13, no 2–3, pp. 9–11.

Beasley J. (1995) Determining teaching and research efficiencies, Journal of the Operational Research Society, №46, pp. 441 – 452.

Bernardo A.B.I. (2003) Towards a Typology of Philippine Higher Education Institutions, Manila: Commission on Higher Education.

Bonaccorsi A., Daraio C. (2009) Characterizing the European university system: a preliminary classification using census microdata, Science and Public Policy №36 (10), pp. 763-775

Boyer E. (1990). Scholarship reconsidered: Priorities of the professorship. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.

Caliński R., Harabasz J. (1974) A dendrite method for cluster analysis. Communications in Statistics, no 3, pp. 1–27.

Charnes A., Cooper W., Rhodes E. (1978) Measuring the efficiency of decision-making units, European Journal of Operational Research, №2, pp. 429 – 444.

Checkoway B. (2001) Renewing the civic mission of the American Research University. The Journal of Higher Education; Mar/Apr 2001; 72, 2; p. 125.

CHEPS (2008) Mapping Diversity: Developing a European Classification of Higher Education Institutions. Enschede, The Netherlands: Center for Higher Education Policy Studies , Center for Higher Education Policy Studies.

Coelli T., Rao D., O'Donnell C., Battese G. (2005) An Introduction to Efficiency and Productivity Analysis, N.Y.: Springer.

Daraio, C., Bonaccorsi, A., Geuna, A., Lepori, B., Bach, L., Bogetoft, P., F. Cardoso, M., Castro-Martinez, E., Crespi, G., de Lucio, I.F., Fried, H., Garcia-Aracil, A., Inzelt, A., Jongbloed, B., Kempkes, G., Llerena, P., Matt, M., Olivares, M., Pohl, C., Raty, T., Rosa, M.J., Sarrico, C.S., Simar, L., Slipersaeter, S., Teixeira, P.N., Eeckaut, P.V. (2011). The European university landscape: A micro characterization based on evidence from the Aquameth project. Research Policy 40, 148–164. doi:10.1016/j.respol.2010.10.009

Education in Figures: Pocket Data Book (2013). – Moscow: National Research University – Higher School of Economics (HSE). – 80 p.

Flexner A. (1994) Universities: American, English, German. New Brunswick, NJ: Transaction Publishers. Originally published: New York: Oxford University Press.

Gokhberg L., Peck M.J., and J. Gacs (eds.) (1997) Russian Applied Research and Development: Its problems and Its Promise. Laxenburg (Austria): IIASA.

Gokhberg L., Kuznetsova T., Zaichenko S. (2011) Russia: Universities in the Context of Reforming the National Innovation System. In: B. Goransson and C. Brundenius (eds.) Universities in Transition: The Changing Role and Challenges for Academic Institutions. Springer.

Humboldt W. Über die innere und äußere Organisation der höheren wissenschaftlichen Anstalten in Berlin. Wilhelm von Humboldts Gesammelte Schriften. Band X: Herausgegeben von der Königlich Preusstschen Akademie der Wissenschaften. Berlin: B. Behr's Verlag, 1903. S. 250-260.

Howells J., Ramlogan R., Cheng S-L. (2008) The Role, Context and Typology of Universities and Higher Education Institutions in Innovation Systems: A UK Perspective (MIoIR Discussion Paper), University of Manchester.

Johnes J. (2006) Data envelopment analysis and its application to the measurement of efficiency in higher education, Economics of Education Review, №25, pp. 273 – 288.

Johnes J., Yu L. (2008) Measuring the research performance of Chinese higher education institutions using data envelopment analysis, China Economic Review, 19, pp. 679–696

Johnes G., Johnes J., Kortelainen M., Thanassoulis E. (2011) Costs and Efficiency of Higher Education Institutions in England: A DEA Analysis, Journal of the Operational Research Society, 62, pp. 1282 – 1297.

Kempkesa G., Pohl C. (2010) The efficiency of German universities – some evidence from nonparametric and parametric methods, Applied Economics, №42, pp. 2063 – 2079.

Marginson S. (2007) University mission and identity for a post post-public era. Higher Education Research & Development, Vol. 26, No. 1, March 2007, pp. 117–131.

McCormick C., Zhao C. (2005) Rethinking and Reframing the Carnegie Classification. Change (September-October), pp. 51–57. Available at: http://classifications.carnegiefoundation.org/downloads/rethinking.pdf (accessed 13 February 2014).

McMillan M., Wing H.C., (2006) University Efficiency: A Comparison and Consolidation of Results from Stochastic and Non-stochastic Methods, Education Economics, №14, pp. 1 – 30.

Mirkin B. (2011) Metody klaster-analiza dlya podderzhki prinyatiya reshenii: obzor [Cluster Analysis for Decision Making: Review]. Preprint WP7/2011/03, Moscow: Publishing House of the National Research University — Higher School of Economics.

Newman H. (1996) Turner, Frank M., ed., The Idea of a University. John Henry Newman (New Haven & London: Yale University Press. Originally published in 1852.

Ortega y Gassett J. (1944). Mission of the university. (H. L. Nostrand, Ed. & Trans.). Princeton, NJ: Princeton University Press.

Scott P. (1995) The Meanings of Mass Higher Education. Buckingham: Open University Press.

Titova N. (ed.) (2008) Strategii razvitiya rossiiskikh vuzov: otvety na novye vyzovy [Development Strategies of Russian Universities: Responses to the New Challenges], Moscow: MAKS Press.

University of Twente (2005) Institutional Profiles Towards a Typology of Higher Education Institutions in Europe. Available at: http://www.utwente.nl/mb/cheps/research/projects/ceihe/publications/socratesceiheinstitutionalprofiles.pdf (accessed 13 February 2014).

Ward J. H., Jr. (1963) Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association, vol. 58, pp. 236–244.

Warning S. (2004) Performance Differences in German Higher Education: Empirical Analysis of Strategic Groups, Review of Industrial Organization, №24, pp. 393 – 408.

Appendix 1. Criteria for classifying Russian universities

Structuring basis (classifier)	Brief description	Examples
By economic model*	Economic model type as a characteristic of the university's diversification and revenues. This group is based on indicators measuring the following revenues (and their ratio): - from education and other activities, - government budget and other public funding provided to finance educational activities	
• Government-supported (most of the funding comes from public sources; low level of diversification)	Revenues from educational activities mainly come from public sources of all kinds; other activities generate a negligible proportion of the university's revenues	Saint Petersburg Mining Institute, Dagestan, Chelyabinsk, and Kaluga State Pedagogical Universities
• Selling their services (educational activities are funded from mixed sources, low level of diversification)	Revenues from educational activities mainly come from non-government sources; other activities generate a negligible proportion of the university's revenues	Moscow State Construction University, Moscow State Law Academy
• Government-supported diversified (educational activities are funded mostly from public sources, high level of diversification)	Revenues from educational activities come from public sources of all kinds; other activities generate a significant proportion of the university's revenues	MISIS, MEPHI, MIPT
• Diversified (educational activities are funded from mixed sources, high level of diversification)	Non-government sources provide a large part of revenues from education activities; other activities generate a significant proportion of the university's revenues	Belgorod State University, Tyumen State Oil and Gas University, Ural State Technical University
• Non-state universities (low level of diversification; not covered by the NRU HSE study)	Revenues are generated mostly from educational activities	Natalia Nesterova Academy, Moscow Academy of Finance and Law
• Non-state universities (high level of diversification; not covered by the NRU HSE study)	Revenues are generated from educational and other activities	New Economic School, Moscow School of Social and Economic Sciences
Strategy employed to adapt	Strategy as a characteristic of	For ethical reasons, names of

Structuring basis (classifier)	Brief description	Examples
to changing environment**	the university's management and/or its way to adapt to the changing environment. This group is based on indicators describing availability of resources to support teaching at the university, financial performance, secondary activities, and quantitative growth (number of students and development of new educational programmes)	universities of various types are not published in open sources. Below we describe profiles of universities most common for the identified types (data for 2005)
Leaders	High basic characteristics	Technics and technology, architecture, construction, creative
Diversifiers	A lot of secondary activities	Finance, economics, law
Expansionists	High quantitative growth (enrolment, introduction of new programmes, etc.)	Architecture, construction, finance, economics
Accumulators (of material, human, and financial resources)	Significant availability of resources	Architecture, construction, finance, economics
Conservatives	Exclusively average characteristics	All types, but mostly classical, humanities, and pedagogics
Outsiders	Below average values	Humanities, pedagogics, classical
Atypicals	Difficult to include into any group	All types

*Research which allowed to classify universities by this criterion was conducted by the NRU HSE's Institute for Educational Studies, covering universities supervised by the Federal Agency for Education [*Abankina et al.*, 2010].

**A number of studies to analyse Russian universities' adaptation strategies were conducted by the NRU HSE researchers [*Titova*, 2008].

Source: [*Abankina et al.*, 2013b]

Appendix 2. Descriptive statistics of parameters for clustering the HEIs

	Min value	Max value	Mean	Standard deviation
R&D	0.00	100.00	58.77	32.44
USE	48.20	90.40	62.69	6.74
Youngsters	0.00	0.46	0.16	0.09
UPF	49.84	352.29	101.07	45.88
Size	303.00	35310.00	8352.37	5793.69

Appendix 3. Descriptive statistics of parameters for evaluation of HEIs performance

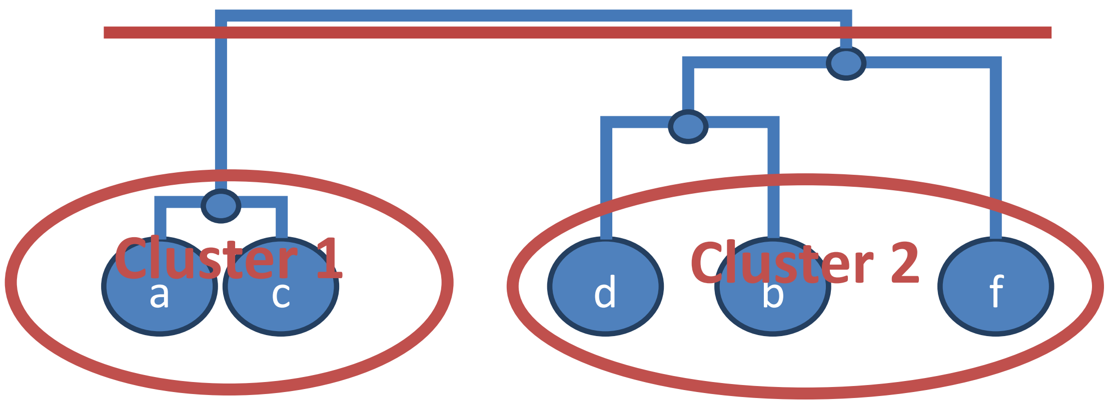
Indicator	Min value	Max value	Mean	Standard deviation
Technical HEIs (tech)				
YPPS	0.02	0.35	0.17	0.07
PUB	2.47	2856.60	135.45	353.36
STUD	1360	19724	8061.78	4416.29
UPF	57.71	252.46	107.99	43.99
PDS	18.21	80.35	63.55	11.04
PPS	39	2914	632.31	519.31
USE	50.10	90.4	60.95	6.32
Classical HEIs (classic)				
YPPS	0.01	0.44	0.18	0.09
PUB	0.79	1338.54	112.48	216.80
STUD	1968	33654	10963.47	6778.37
UPF	57.92	266.70	104.40	44.14
PDS	34.61	87.06	64.42	10.85
PPS	26	2233	518.60	347.99
USE	51.80	76.90	63.45	5.97
Pedagogical HEIs (ped)				
YPPS	0.06	0.38	0.18	0.07
PUB	2.13	1645.58	124.43	313.40
STUD	926	16318	5293.93	3288.26
UPF	50.99	126.46	75.82	18.31
PDS	37.90	81.69	64.60	7.77
PPS	71	1610	636.89	367.92
USE	51.90	68.80	59.80	3.73
Socio-economic HEIs (soc_econ)				
YPPS	0.03	0.46	0.19	0.10
PUB	1.48	2260.78	304.78	631.52
STUD	2278	35310	10041.92	7942.90
UPF	49.83	164.81	84.02	27.51

Indicator	Min value	Max value	Mean	Standard deviation
PDS	45.20	83.80	65.98	10.23
PPS	27	4136	646.64	895.59
USE	56.7	82.6	68.36	6.65
Arcitechture HEIs (arc)				
YPPS	0.02	0.26	0.16	0.07
PUB	3.98	444.44	137.32	152.91
STUD	1269	11642	5229.50	2970.45
UPF	59.67	174.45	94.91	41.96
PDS	57.02	73.83	67.75	4.82
PPS	42	3166	824.5	891.10
USE	51.80	76.90	64.26	6.81
Humanitarian HEIs (hum)				
YPPS	0.02	0.24	0.14	0.06
PUB	1.27	739.21	175.92	232.56
STUD	303	21602	5468.63	6334.34
UPF	73.24	352.29	137.30	92.07
PDS	57.06	76.97	68.49	7.08
PPS	117	634	353.63	152.55
USE	50.5	85.7	65.97	9.81
All HEIs				
YPPS	0.01	0.46	0.18	0.08
PUB	0.79	2856.60	151.41	355.16
STUD	303	35310	8396.02	5931.52
UPF	49.83	352.29	100.26	45.48
PDS	18.22	87.06	64.74	10.04
PPS	26	4136	598.58	537.97
USE	50.1	90.40	62.83	6.71

Appendix 4. Description of hierarchical clustering

As a first step of hierarchical clustering, a couple of objects less different from each other than all the rest in the set, are added to a cluster (see Figure 1 below).

Figure 1. Visualisation of hierarchical clustering



Source: composed by the authors.

At the next iteration, two closest objects (or previously created clusters) are merged together (depending on where the difference metric has a smaller value)

$$K_{ij}^{(t)} = K_i^{(t-1)} \cup K_j^{(t-1)}, \text{ if } d\left(K_i^{(t-1)}, K_j^{(t-1)}\right) = \min_{s,r} d(K_s^{(t-1)}, K_r^{(t-1)}) \quad (1)$$

Nothing changes for all other clusters, i. e.,

$$K_l^{(t)} = K_l^{(t-1)} \quad (2)$$

These steps are repeated until all clusters are combined into one. The end result of applying this algorithm is a tree (or cluster hierarchy).

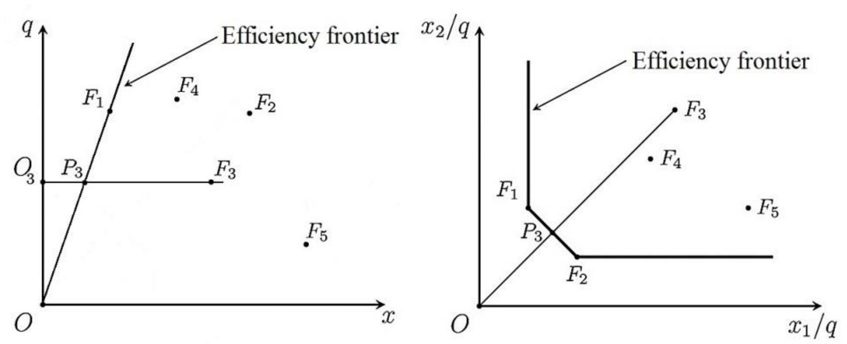
Appendix 5.Visualisation of DEA model: the case of 1-1 and 1-2 inputs-outputs set

Upon assumption of constant return to scale, a HEI exploits one input *x* to produce only one output *q* (Figure 1.1). The efficiency score for a HEI at the point F_3 can be calculated as the ratio of O_3P_3/O_3F_3 . It is worth noting that F_1 is the single efficient HEI as being a point on the efficient frontier and having $\theta = 1$.

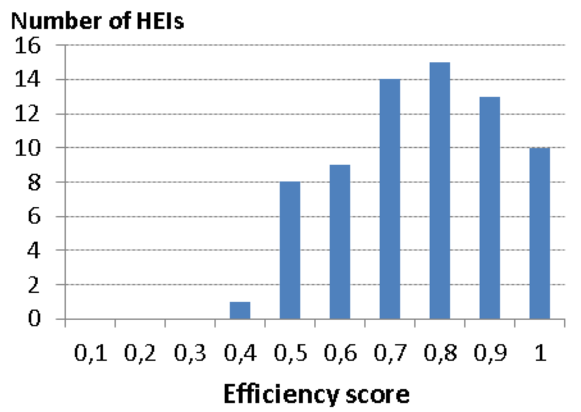
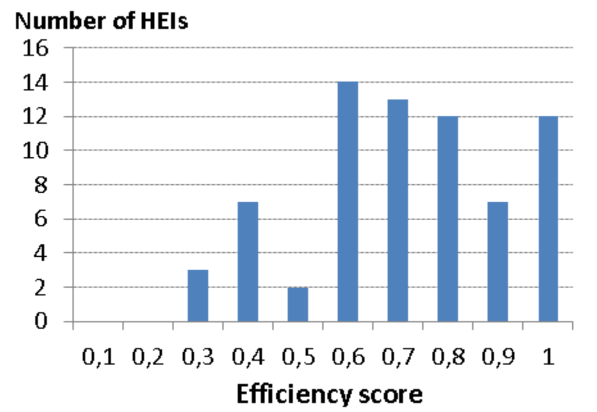
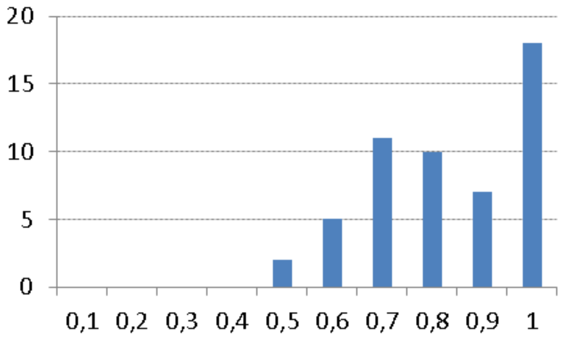
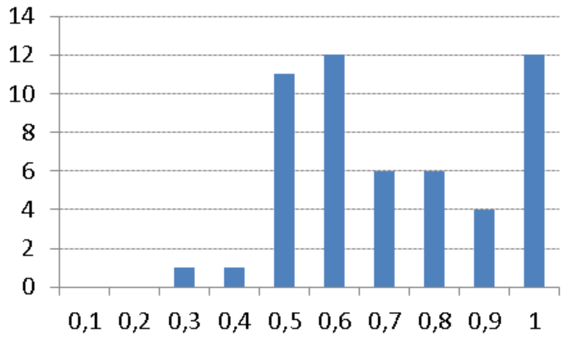
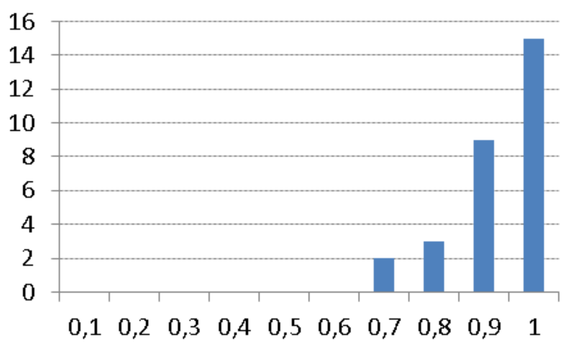
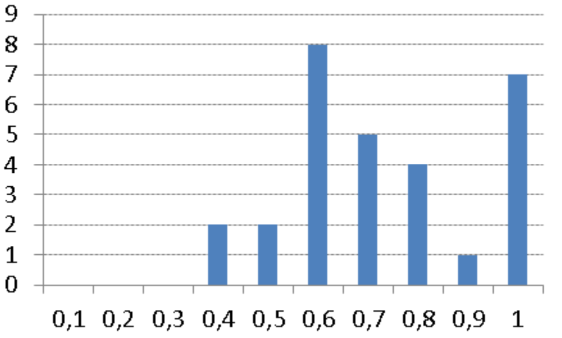
Given the example (Figure 1.2) of two inputs ($x_1$and x_2) and one output (*q)*,the efficient envelopment is constructed from two efficient HEIs: F_1 and F_2, respectively. They become peers for the F_3 unit. Thereafter, the efficiency of F_3 could be measured by the ratio of OP_3/O_3F_3.

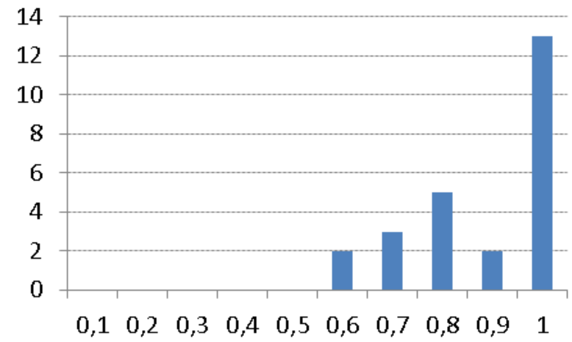
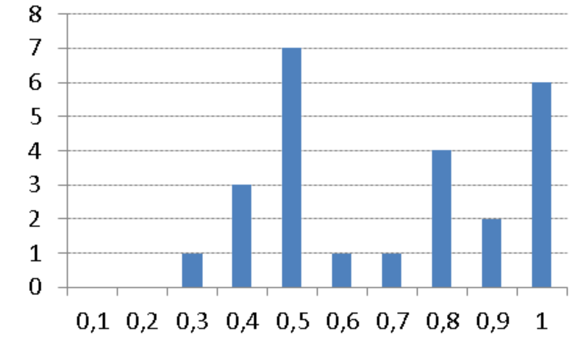
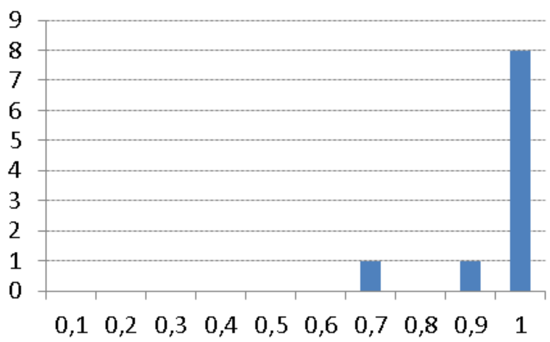
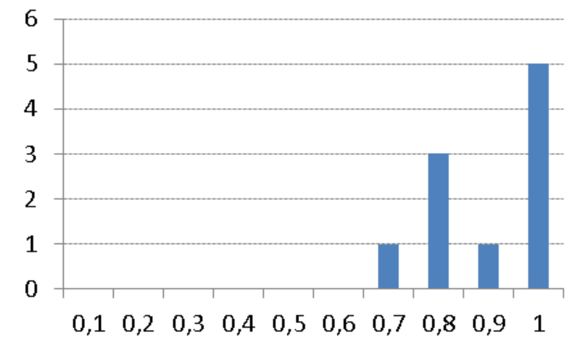
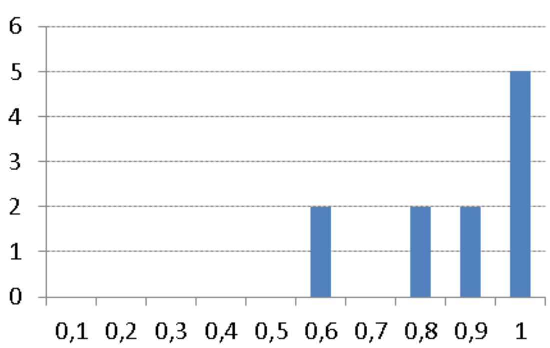
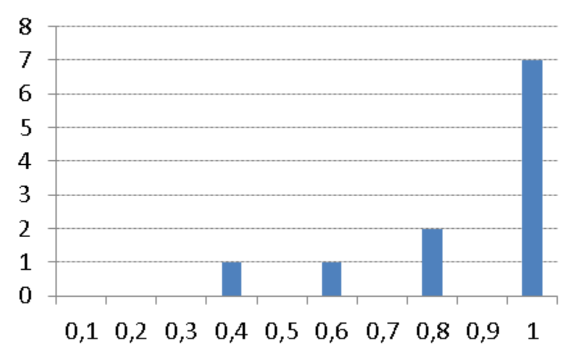
Figure 1. DEA model in the case of 1-1 and 1-2 inputs-outputs set

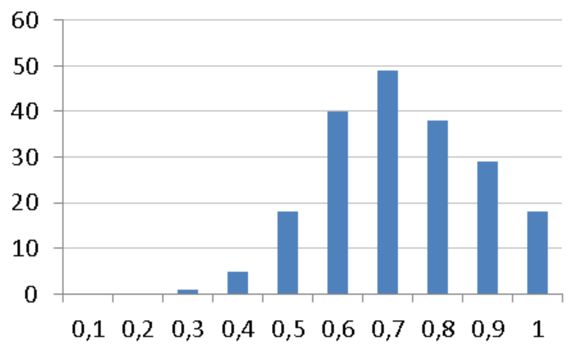
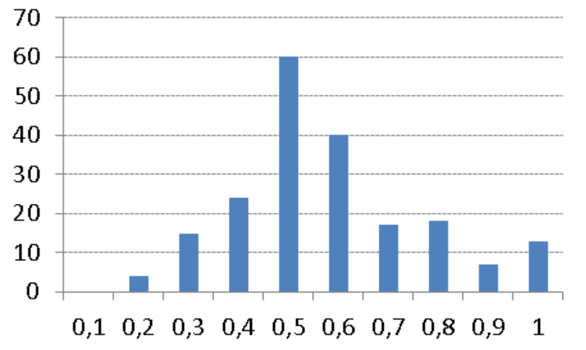
(*x* is an input, *q* is an output).



Appendix 6. Distribution of efficiency scores for homogeneous groups of HEIs, model 1 and 2

Model 1	Model 2
Technical HEIs	
	
Classical HEIs	
	
Pedagogical HEIs	
	

<table>
<tr><th>Model 1</th><th>Model 2</th></tr>
<tr><td colspan="2">Socio-economic HEIs</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">Architectural HEIs</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">Humanities HEIs</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">All HEIs</td></tr>
</table>

Model 1	Model 2
	

Appendix 7. Descriptive statistics for homogeneous groups of HEIs, model 1 and 2

	Model 1	Model 2
Technical HEIs		
Average score (%)	71.43	66.91
Standard deviation (%)	16.29	21.05
Classical HEIs		
Average score (%)	78.93	67.67
Standard deviation (%)	16.42	20.87
Pedagogical HEIs		
Average score (%)	90.14	70.1
Standard deviation (%)	10.43	21.27
Socio-economic HEIs		
Average score (%)	84.76	65.78
Standard deviation (%)	15.57	24.91
Architectural HEIs		
Average score (%)	94.31	87.05
Standard deviation (%)	9.77	14.68
Humanities HEIs		
Average score (%)	84.06	85.40
Standard deviation (%)	17.89	20.54
All HEIs		
Average score (%)	68.18	53.21
Standard deviation (%)	15.74	19.17

Appendix 8.Robustness Check of DEA results

Table 1. Pearson correlation matrix

Model 1 — *Technical HEIs(tech)*

tech	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9453	0.7199	0.5569*
m = 0.2		1	0.8323	0.6127
m= 0.5			1	0.6096
m=0.8				1

Model 2 — *Technical HEIs(tech)*

tech	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9857	0.9004	0.8224*
m = 0.2		1	0.9462	0.8572
m= 0.5			1	0.9169
m=0.8				1

Classical HEIs (classic) — Model 1

classic	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.8828	0.8492	0.7089
m = 0.2		1	0.6511	0.5885
m= 0.5			1	0.5642*
m=0.8				1

Classical HEIs (classic) — Model 2

classic	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9866	0.9030	0.6763*
m = 0.2		1	0.9336	0.7015
m= 0.5			1	0.7751
m=0.8				1

Pedagogical HEIs (ped) — Model 1

ped	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9834	0.8865	0.8441*
m = 0.2		1	0.9383	0.8778
m= 0.5			1	0.8772
m=0.8				1

Pedagogical HEIs (ped) — Model 2

ped	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9919	0.9256	0.7545*
m = 0.2		1	0.9590	0.7927
m= 0.5			1	0.8515
m=0.8				1

Socio-economic HEIs (soc_econ) — Model 1

soc_econ	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9937	0.9795	0.8160**
m = 0.2		1	0.9845	0.8077*
m= 0.5			1	0.8674
m=0.8				1

Socio-economic HEIs (soc_econ) — Model 2

soc_econ	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9877	0.9493	0.8537*
m = 0.2		1	0.9757	0.8753*
m= 0.5			1	0.9206
m=0.8				1

Architectural HEIs (arch) — Model 1

arch	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9905	0.9709	0.7083
m = 0.2		1	0.9823	0.6770
m= 0.5			1	0.7797
m=0.8				1

Architectural HEIs (arch) — Model 2

arch	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9433	0.7434	0.9083
m = 0.2		1	0.8722	0.7890
m= 0.5			1	0.6132**
m=0.8				1

Model 1					Model 2				
HumanitiesHEIs (hum)									
hum	DEA	m = 0.2	m= 0.5	m=0.8	hum	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9733	0.9016*	0.9165**	DEA	1	0.9792	0.9737	0.3104*
m = 0.2		1	0.9265	0.9721	m = 0.2		1	0.9849	0.1484*
m= 0.5			1	0.9696	m= 0.5			1	0.0938*
m=0.8				1	m=0.8				1
All HEIs									
all	DEA	m = 0.2	m= 0.5	m=0.8	all	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9278	0.7345	0.5111	DEA	1	0.9856	0.8994	0.6747*
m = 0.2		1	0.7618	0.4217*	m = 0.2		1	0.9396	0.6986
m= 0.5			1	0.5267	m= 0.5			1	0.8015
m=0.8				1	m=0.8				1

Note: * not statistically significant on 10% level of confidence,
** not statistically significant on 5% level of confidence

Table 2. Spearman correlation matrix

Model 1					Model 2				
Technical HEIs(tech)									
tech	DEA	m = 0.2	m= 0.5	m=0.8	tech	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9538	0.6920	0.5145	DEA	1	0.9943	0.9278	0.8230
m = 0.2		1	0.7488	0.5420	m = 0.2		1	0.9392	0.8155
m= 0.5			1	0.4550*	m= 0.5			1	0.7896*
m=0.8				1	m=0.8				1
Classical HEIs (classic)									
classic	DEA	m = 0.2	m= 0.5	m=0.8	classic	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9405	0.8981	0.7808	DEA	1	0.9919	0.9472	0.7694
m = 0.2		1	0.8392	0.7372	m = 0.2		1	0.9572	0.7701
m= 0.5			1	0.6560*	m= 0.5			1	0.7401*
m=0.8				1	m=0.8				1
Pedagogical HEIs (ped)									
ped	DEA	m = 0.2	m= 0.5	m=0.8	ped	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9943	0.9141	0.8384	DEA	1	0.9960	0.9722	0.8678
m = 0.2		1	0.9155	0.8451	m = 0.2		1	0.9796	0.8693
m= 0.5			1	0.7777*	m= 0.5			1	0.8544*
m=0.8				1	m=0.8				1
Socio-economic HEIs (soc_econ)									
soc_econ	DEA	m = 0.2	m= 0.5	m=0.8	soc_econ	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9976	0.9817	0.8919*	DEA	1	0.9953	0.9837	0.9233*
m = 0.2		1	0.9793	0.8879*	m = 0.2		1	0.9822	0.9264*

Model 1					Model 2				
m= 0.5			1	0.9229	m= 0.5			1	0.9240*
m=0.8				1	m=0.8				1

Architectural HEIs (arch)

arch	DEA	m = 0.2	m= 0.5	m=0.8	arch	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1.0000	1.0000	0.9846	0.9231*	DEA	1	0.9862	0.9310**	0.9586
m = 0.2		1.0000	0.9846	0.9231*	m = 0.2		1	0.9586	0.9172*
m= 0.5			1.0000	0.9692	m= 0.5			1	0.8759*
m=0.8				1.0000	m=0.8				1

Humanities HEIs (hum)

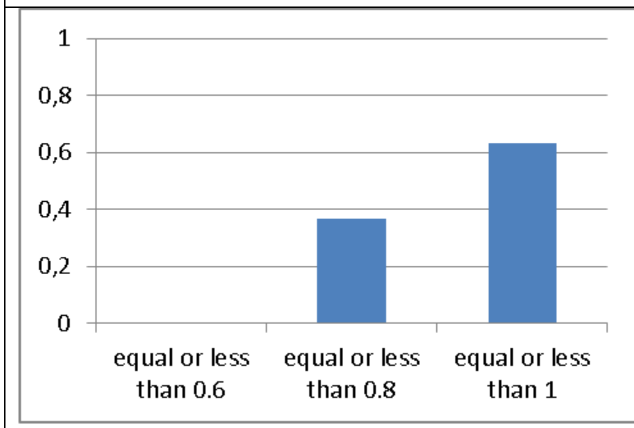
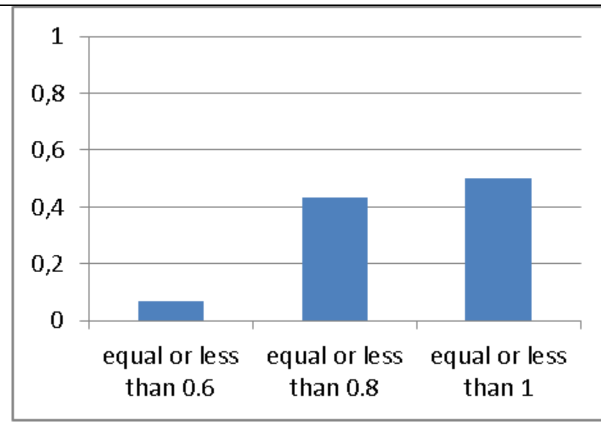
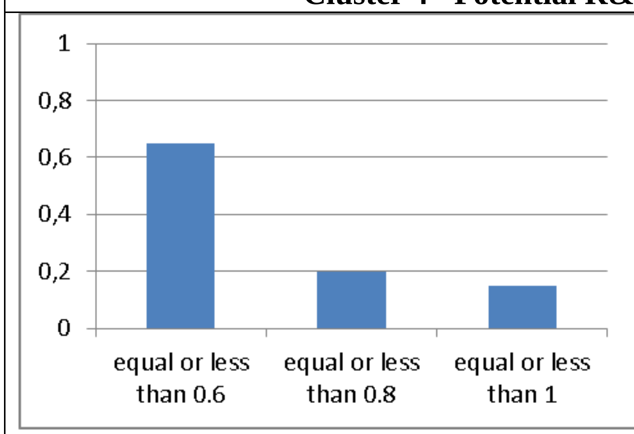
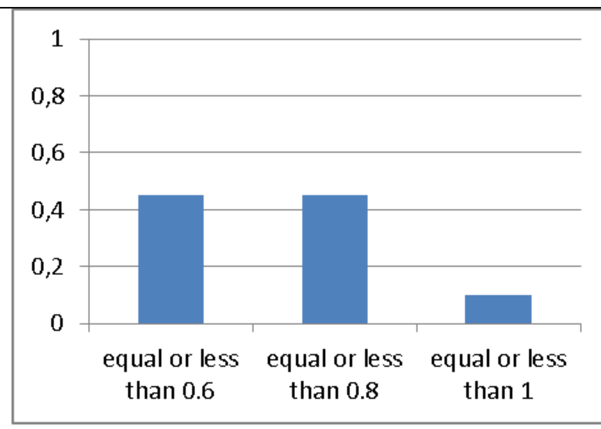
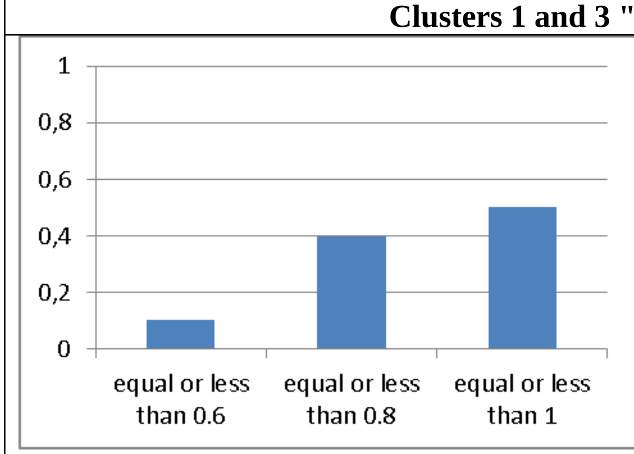
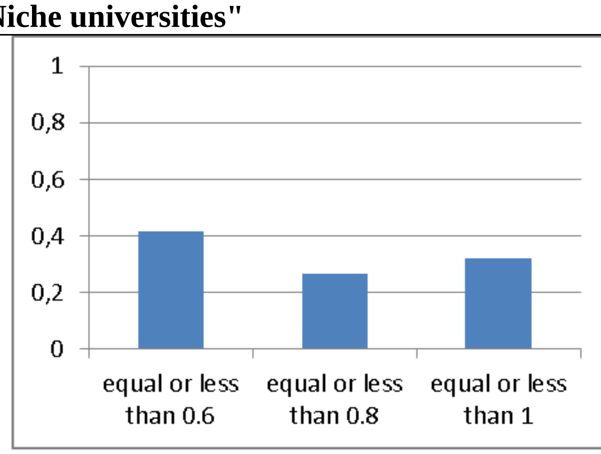
hum	DEA	m = 0.2	m= 0.5	m=0.8	hum	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9907	0.9537	0.9722	DEA	1	1.0000	0.9900	0.7800
m = 0.2		1	0.9352	0.9907	m = 0.2		1	0.9900	0.7800
m= 0.5			1	0.9074*	m= 0.5			1	0.7600
m=0.8				1	m=0.8				1

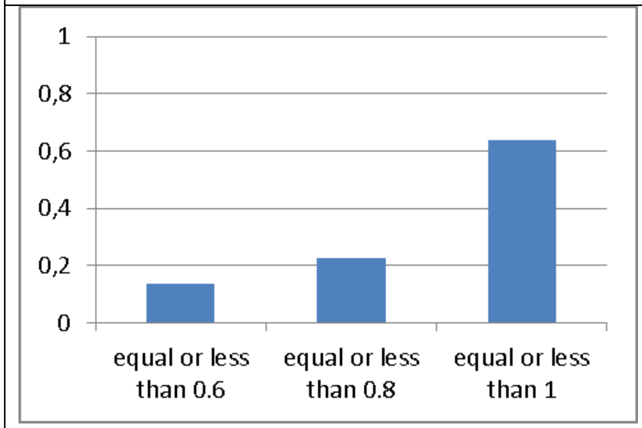
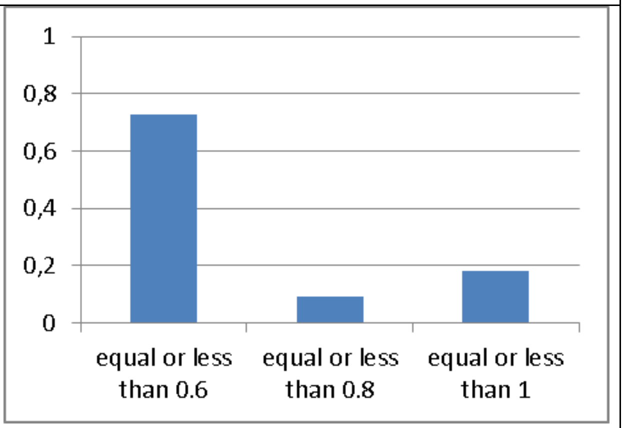
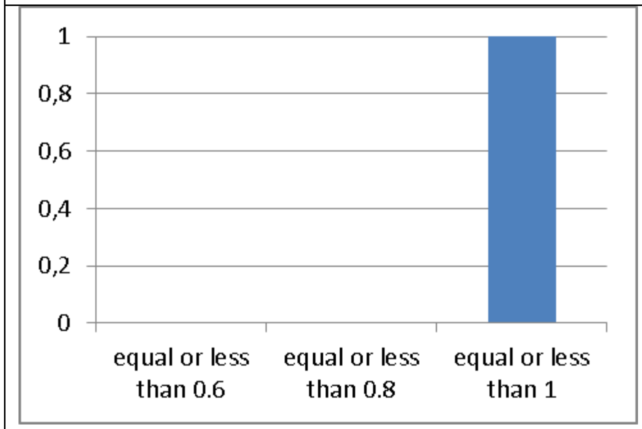
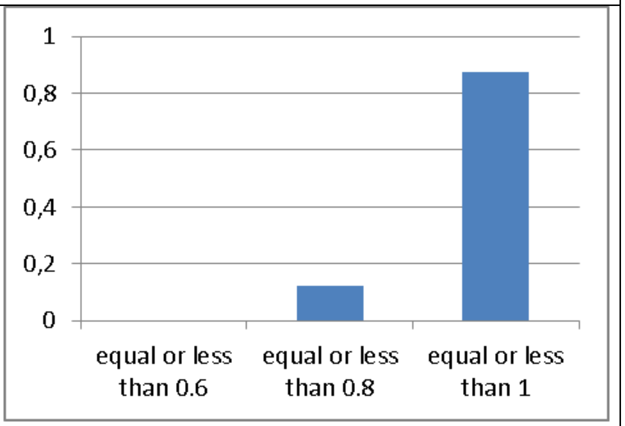
All HEIs

all	DEA	m = 0.2	m= 0.5	m=0.8	all	DEA	m = 0.2	m= 0.5	m=0.8
DEA	1	0.9327	0.7211	0.5012	DEA	1	0.9930	0.9511	0.6605*
m = 0.2		1	0.7328	0.4305*	m = 0.2		1	0.9721	0.6639*
m= 0.5			1	0.5156	m= 0.5			1	0.6862
m=0.8				1	m=0.8				1

Note: * not statistically significant on 10% level of confidence,
** not statistically significant on 5% level of confidence

Appendix 9.Distribution of universities performance by clusters *(in points)*

<table>
<tr><th>Model 1</th><th>Model 2</th></tr>
<tr><td colspan="2">Cluster 2
"Market leaders"</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">Cluster 4 "Potential R&D and education leaders"</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">Clusters 1 and 3 "Niche universities"</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Model 1	Model 2
Cluster 5 "Indeterminate position"	
	
Cluster 6 "Universities of good standing"	
	

Veronika Belousova
Department for Methodology of Budget Planning, Institute for Statistical Studies and Economics of Knowledge.
National Research University - Higher School of Economics, E-mail: vbelousova@hse.ru
20, Myasnitskaya str., Moscow, 101000, Russian Federation