Attractivnes of Programs for Students, Heterogeneity of Higher Educational System, Efficiency of Universities

Fuad Aleskerov

NRU Higher School of Economics

and Institute of Control Sciences

Pixel Conference, Florence, Italy

21-22.03.2019

How can we evaluate
a heterogeneity
of higher education system?

Co-authors

Aleskerov F., Carnoy M., Frumin, I., Ivanova A., Kardanova E., Myachin A., Serova A., Yakuba V.

Higher School of Economics, Moscow, Russia,

Stanford University,

Russian Academy of Sciences

Studying heterogeneity in education

Different views

- ✓ diversification of higher education institutions (*Reichert, 2009; Carnevale; Strohl, 2010; Posselt et al., 2012)*
- ✓ selectivity of higher education institutions, (*Calmand et al., 2009; Hurwitz, 2011; Pastine & Pastine, 2012)*
- ✓ heterogeneity of the student population (*van Ewijk, 2010; De Paola, Scoppa, 2010; Bielinska-Kwapisz and Brown, 2012*)

Different methods

- ✓ statistical and econometric tools
  - Standard deviation, coefficient of variation (*Murdoch, 2002*)
  - Gini coefficient (*Bosi, Seegmuller, 2006; Sudhir, Segal, 2008*)
  - Multidigraphs (degree of hirerarchy , *Fedriani and Moyano, 2011*)

About our work

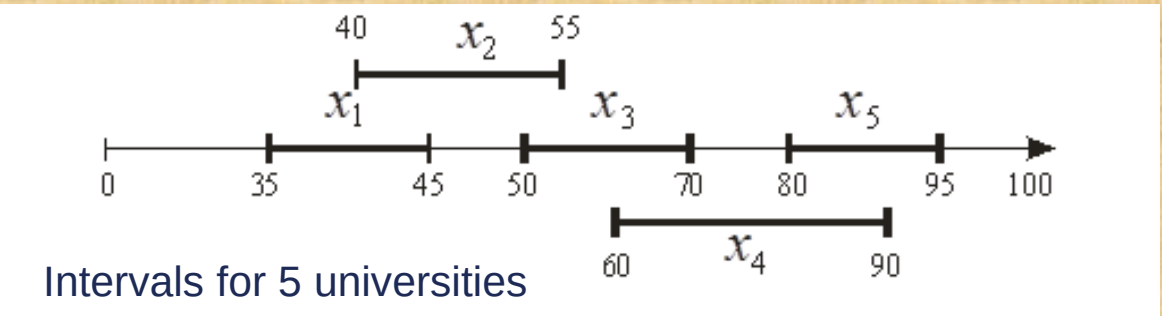
- Another approach to estimate **heterogeneity** in education
- Heterogeneity of an **higher educational system**
- A mathematical model based on the construction of universities' **interval order**
- *The Unified State Examination (USE) scores of Russian students are used to illustrate how our measure of the system's heterogeneity works*.

Our model:
Construction of the interval order

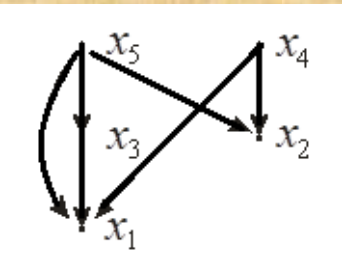
The set of all universities A

For each university i it is known the set of students entrance grades

- Assume that each university is defined by some interval $[a_1; b_1]$ depending on the entrance exam grade.
- Use the mean value of grades and standard deviation to construct the interval order
- If the left boundary for i lies at the right of the right boundary of j, then we include the pair (i,j) to the interval order



Intervals for 5 universities



Graph of the interval order for the 5 universities

$$P = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

The incidence matrix

Another approach to evaluate the heterogeneity

1. The interval order P is constructed
2. The notion of ideal interval order P_{id} is defined

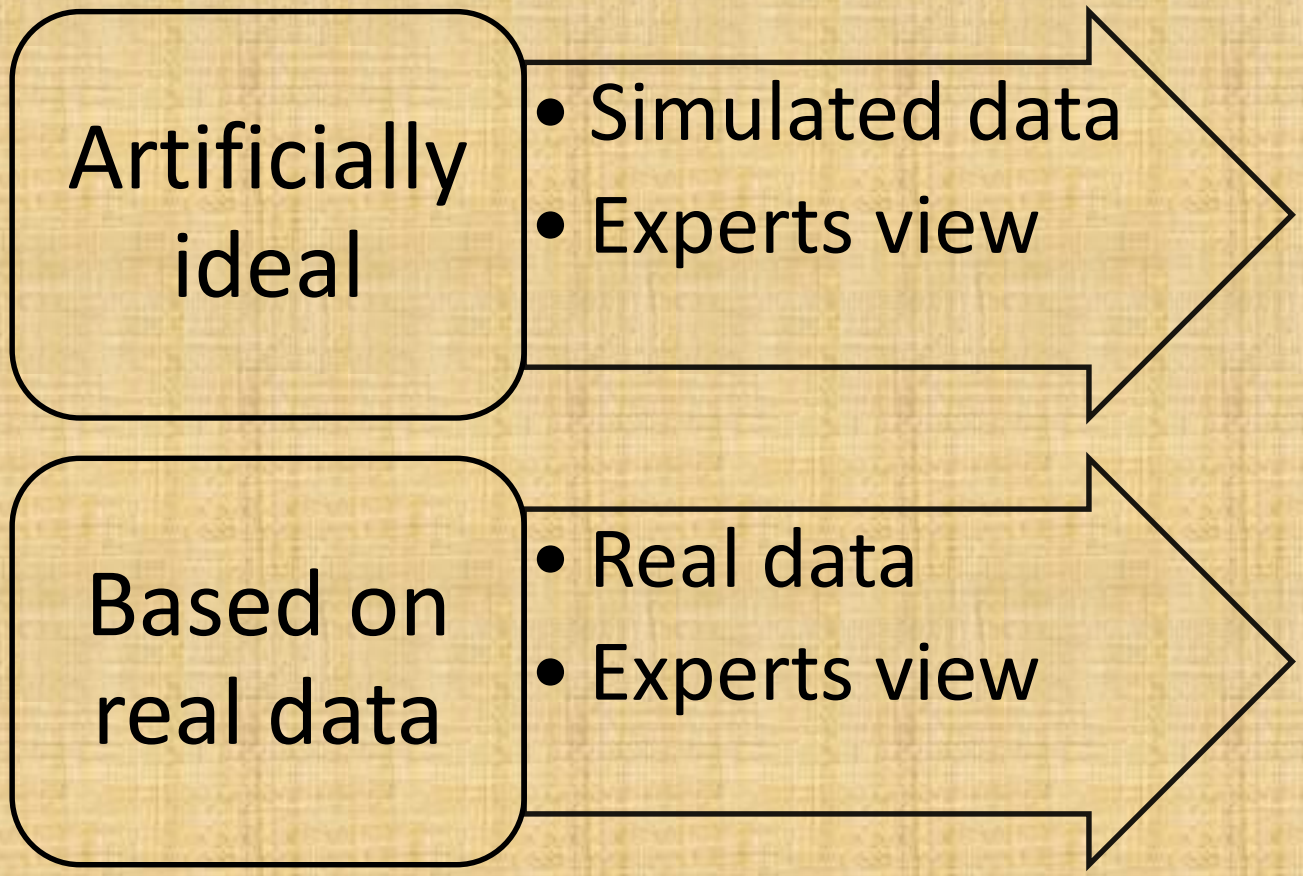
Then as the measure of heterogeneity the Hamming distance is used.

$$H(B_1, B_2) = \frac{1}{n(n-1)} \sum_{i,j} |b^1_{ij} - b^2_{ij}|,$$

where b^1_{ij} is equal to 1 if the pair (i ; j) belongs to the interval order B_1, otherwise =0; b^2_{ij} is equal to 1 if the pair belongs to the interval order B_2 , otherwise it is equal to 0.

Comparing the matrices for the real and the ideal interval orders and using Formula we calculate the *Hamming distance* between two interval orders

The notion of ideal system



Our ideal system

Ideal educational system for Economics and Management: expert view

- ✓ 1) A group of best universities that train managers, strategists, high-class analysts (about 10% of universities, the average score of the whole contingent of enrolled students should not fall below 75)
- ✓ 2) A group of strong universities that train strong professionals for regional labor markets (about 70% of universities, the average score from 65 to 74).
- ✓ 3) Group of universities preparing bachelors on applied programs (about 20% of universities, an average USE score of the admitted contingent should not be lower than 55).

The results

Prototype for ideal system.

The scores' interval	Mean	St.Dev.	Count (%)
>75	79	2.82	10 (3%)
(65;75]	69	3.29	52 (14%)
(55;65]	59	2.67	220 (58%)
<=55	52	1.26	97 (25%)

Comparing the matrices P for real and ideal interval orders and using formula (2) we can calculate the Hamming distance between two interval orders: $H(P, P_{id})=0.26$.

Improving system

The desirable lower limit of the average USE scores for economics majors lies at the level of 55 points

97 universities to delete

The Hamming distance between real and ideal interval orders becomes $H(P,P_{id})= 0.16.$

To conclude

- A new method of studying heterogeneity in the higher education system
- Our method is based on the comparison of the hypothetical educational system to the real system
- We showed how our method works on Russian data
- The model proposed can be applied for any other data, educational systems, countries

How to forecast the demand for higher education ?

- Coauthors

Irina Abankina, Tatyana Abankina, Fuad Aleskerov, Kirill Zinkovsky, Dmitriy Ogorodniychuk, Ludmila Filatova, Pavel Derkachev, Elena Nikolaenko, Eduard Seroshtan, Ludmila Egorova

Proposed system of enrollment forecast in Russian universities

Forecast of dynamics in all sectors of economy by the Ministry of Economy

Study of the needs of companies and investment programs

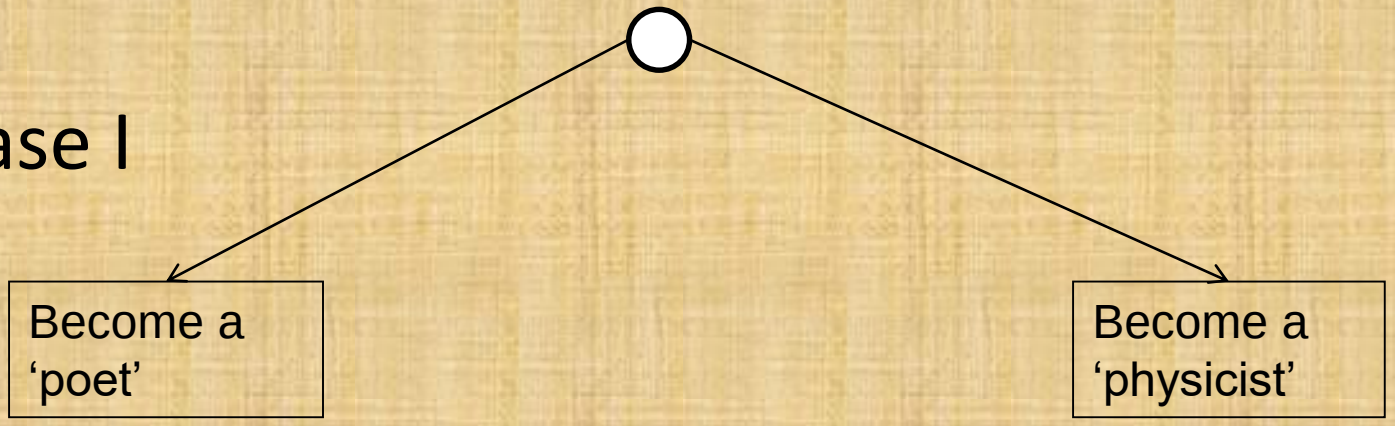
Forecast of the demand of population on the basis of programs' attractiveness, household preferences, priorities of the state and investors, and the trends of economic development, scientific and innovative activities

Forecast of the number of enrolled students by professional programs

photo

Binary choice of entrants

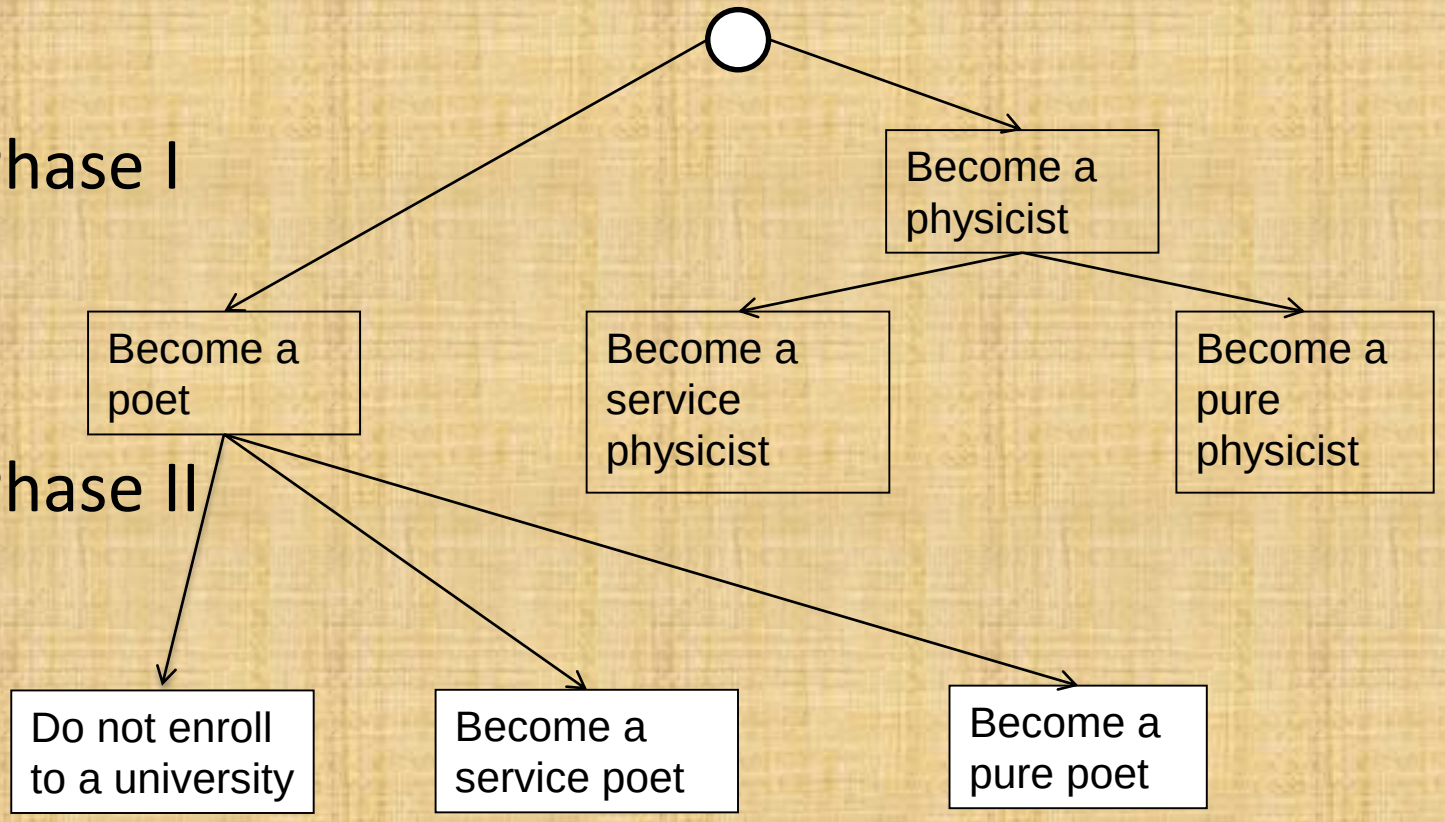
- Phase I



	Aggregated groups of professions	
1	Math and natural sciences	physicists
2	Engineering, technology and engineering sciences	physicists
3	Health and Medical Science	physicists
4	Agricultural sciences	physicists
5	Social sciences	poet
6	Educational Sciences	poet
7	Humanities	poet
8	Art and Culture	poet

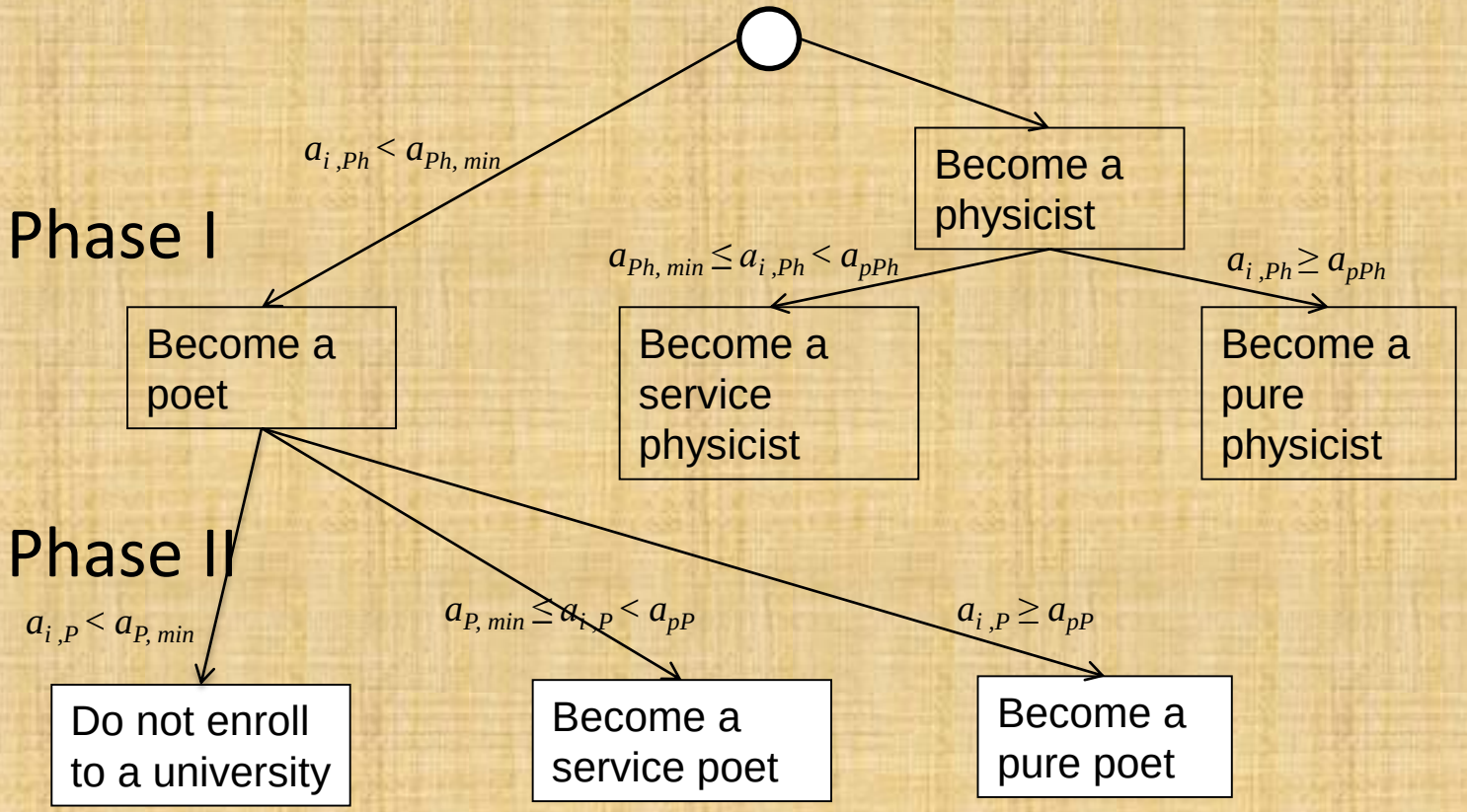
Next steps of entrants' selection

- Phase I
- Phase II



A model

- Phase I
- Phase II



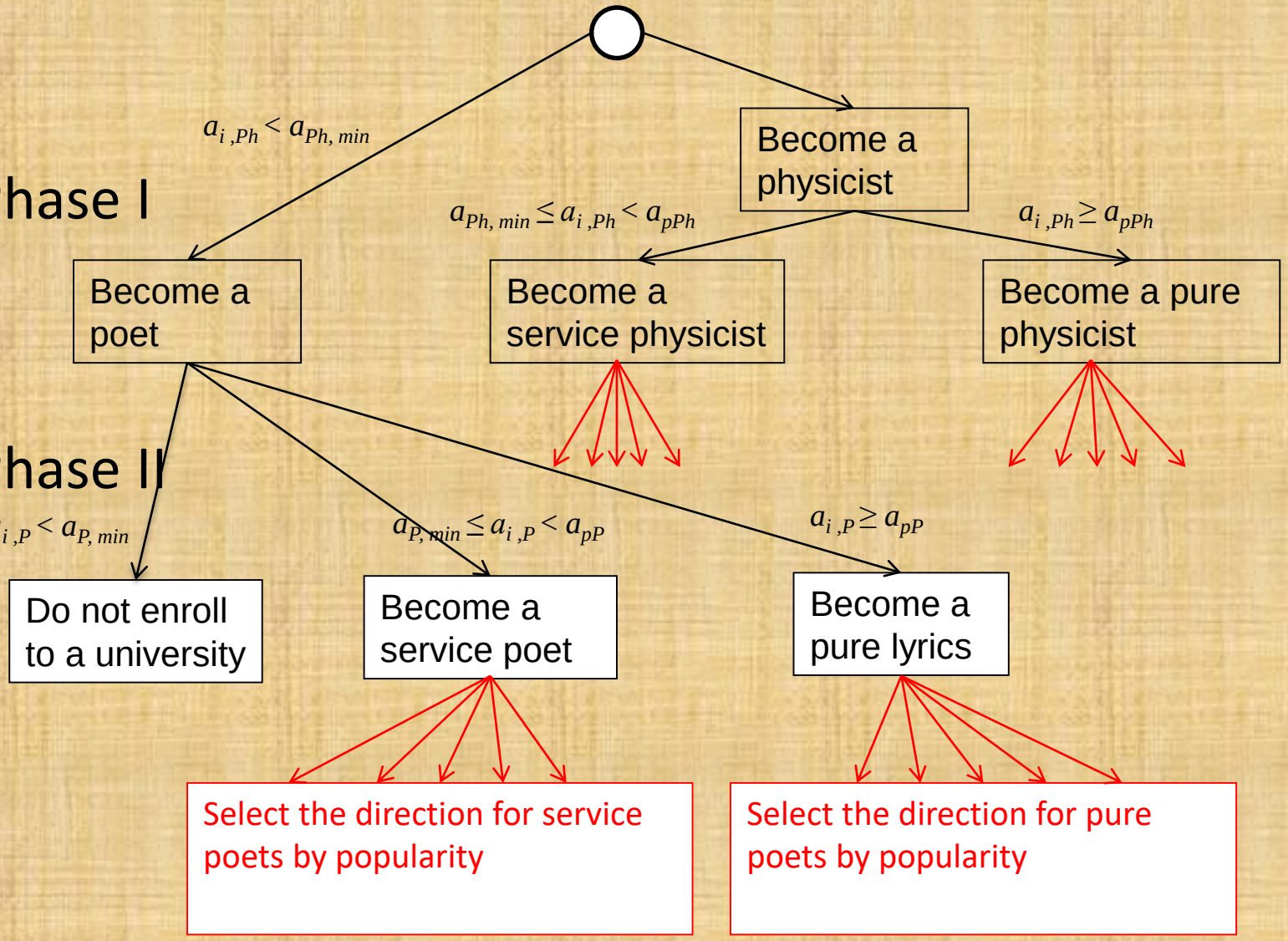
a_{pPh}, a_{pP} – threshold values of Unified State Exam scores for pure physicists and poets,

$a_{Ph,min}$, $a_{P,min}$ – minimum passing Unified State Exam scores for the for pure physicists and poets,

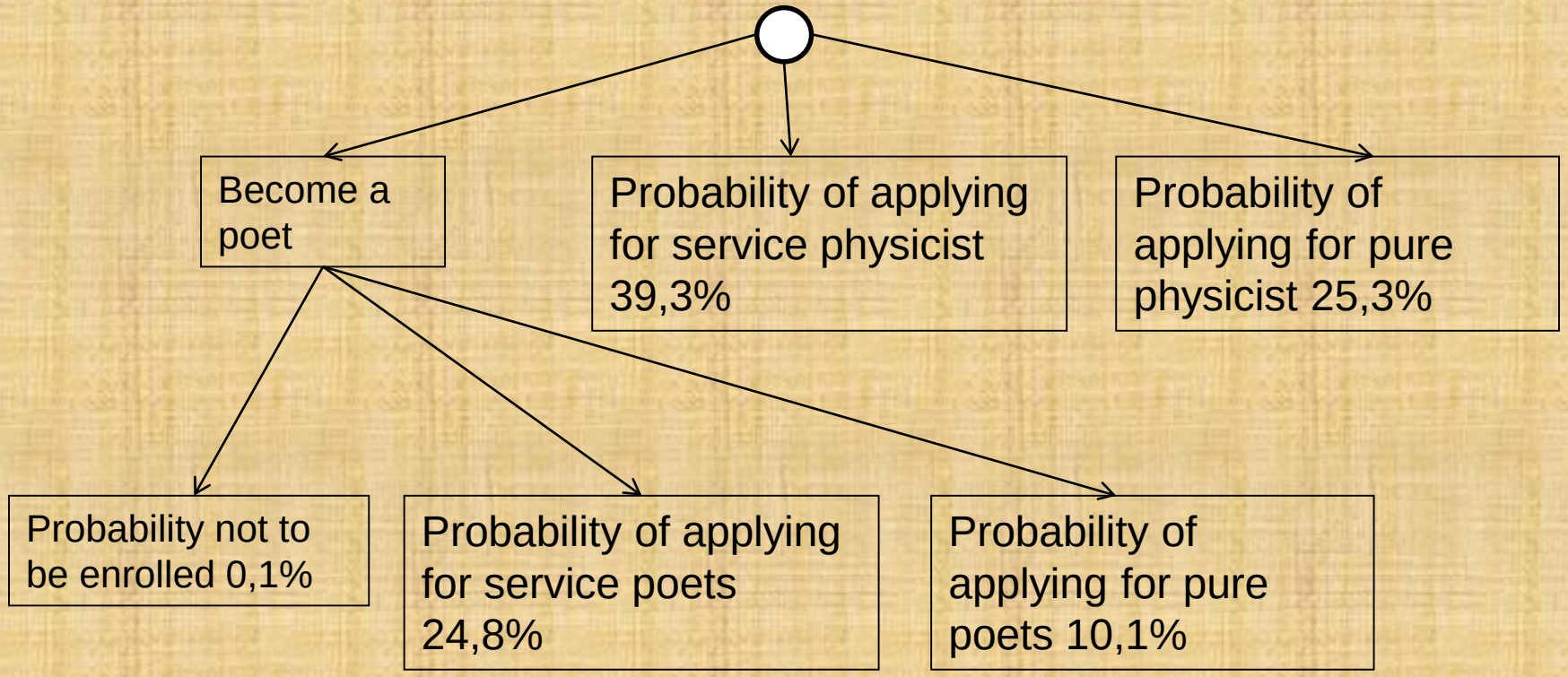
$a_{i,j}$ – ability of entrant *i* to study on the specialization *j*

An elaborated model

- Phase I
- Phase II



Evaluations



Evaluation of professions' attractiveness by methods of semantic analysis

Factors for semantic analysis

1. Attractiveness of the life style and the image of man in this profession that is transmitted through the media
2. Prospect of this profession, broadcasted through the media
3. Quality of education in this profession
4. Accessibility of higher education this profession
5. Interest of society and the popularity of the profession, reflected in the media
6. Attention of the state, corporations and investors to the profession, the direction of training
7. Scientific activity in the treated area, innovation, activity experts
8. Government policies and programs

photo

photo

Characteristic of media sources for semantic analysis

• more than 88 million articles of the Russian media;

• More than 10,000 media: newspapers, magazines, news agencies, Internet publications, television and radio stations;

• The media in all regions of Russia and CIS countries, as well as more than 10 foreign countries;

• more than 70 thousand new documents every day;

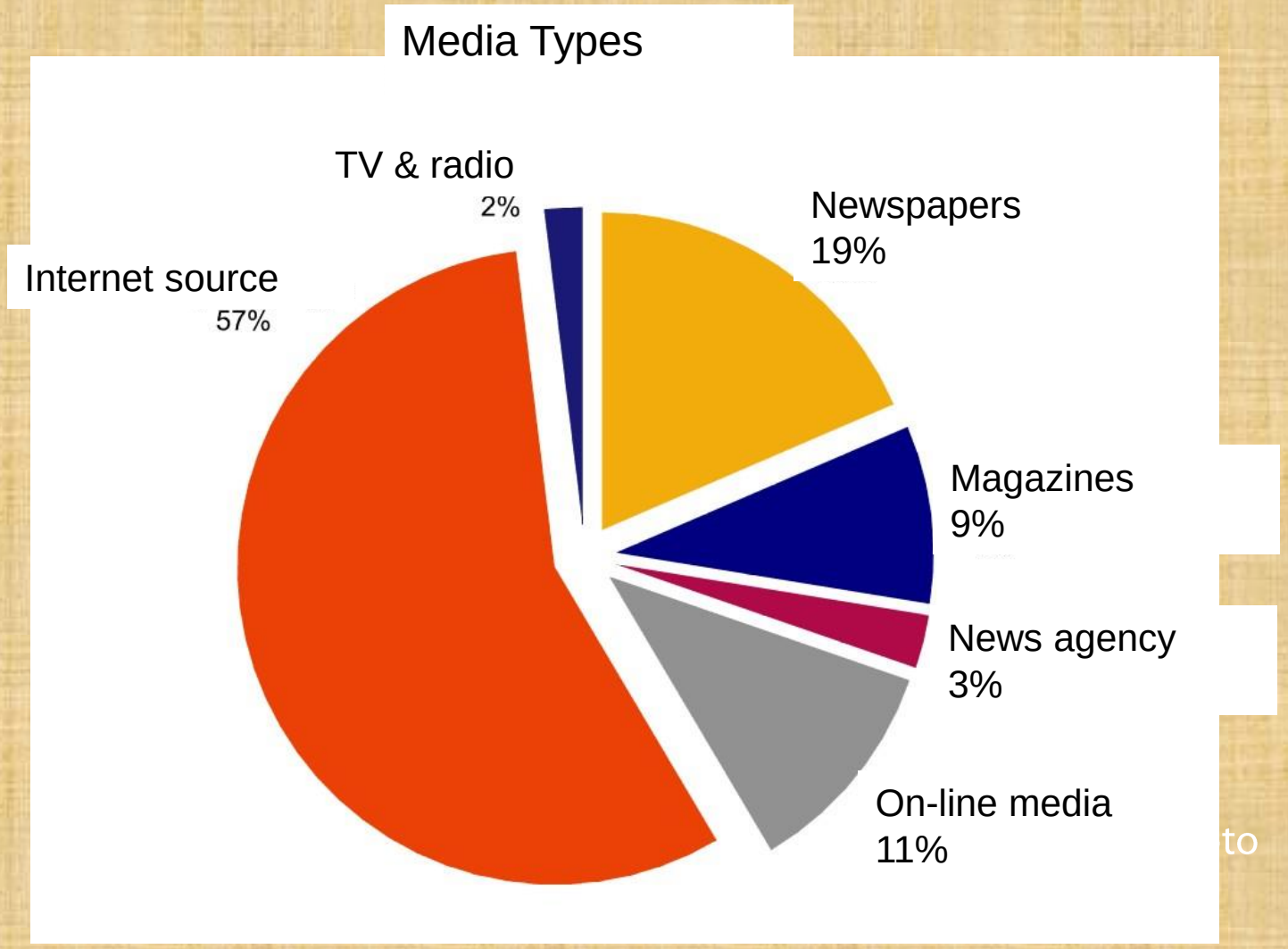
photo

• 100% compliance with the electronic versions of publications released by major newspapers and magazines;

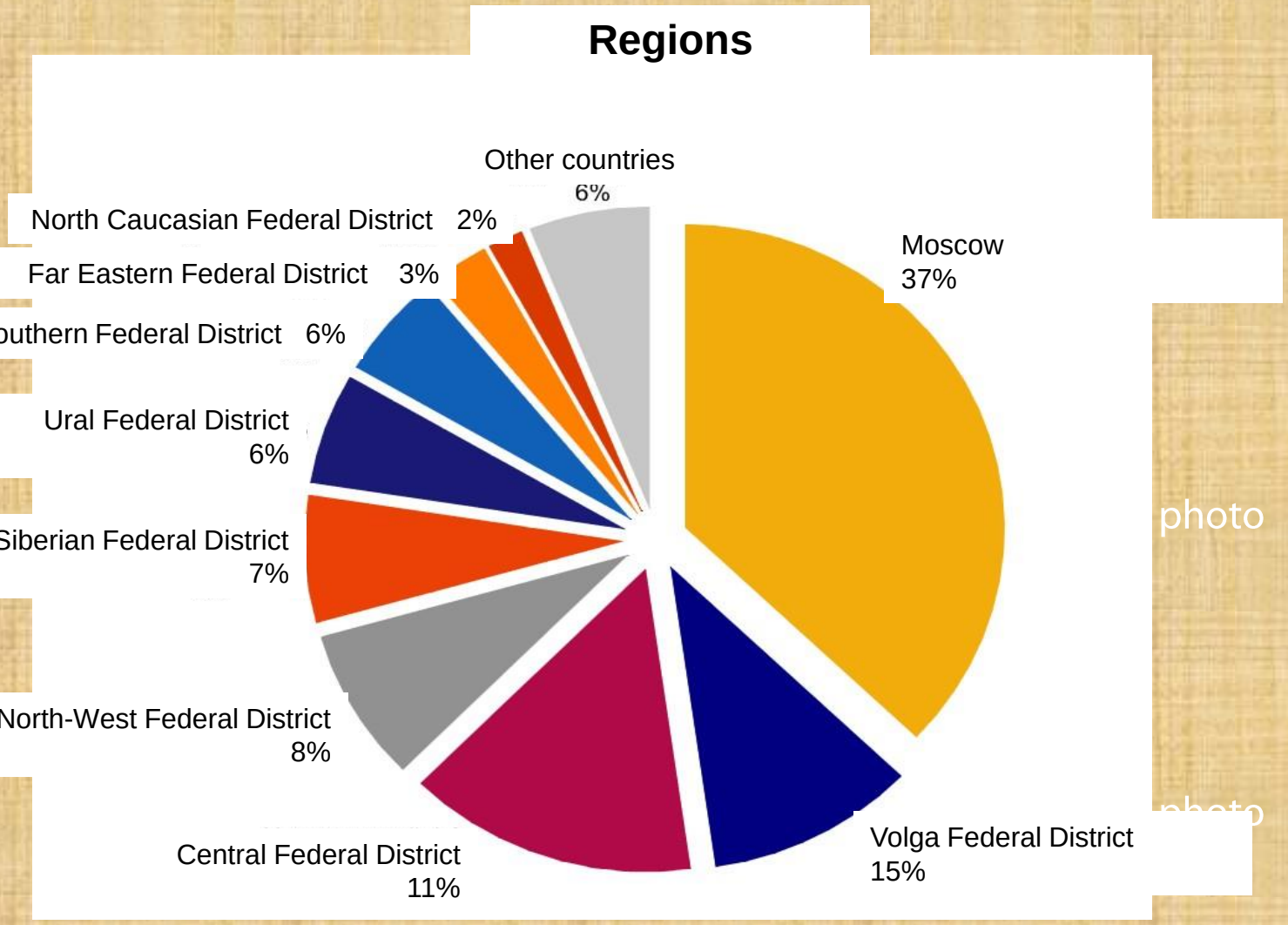
• archival materials of Russian media since 1990

photo

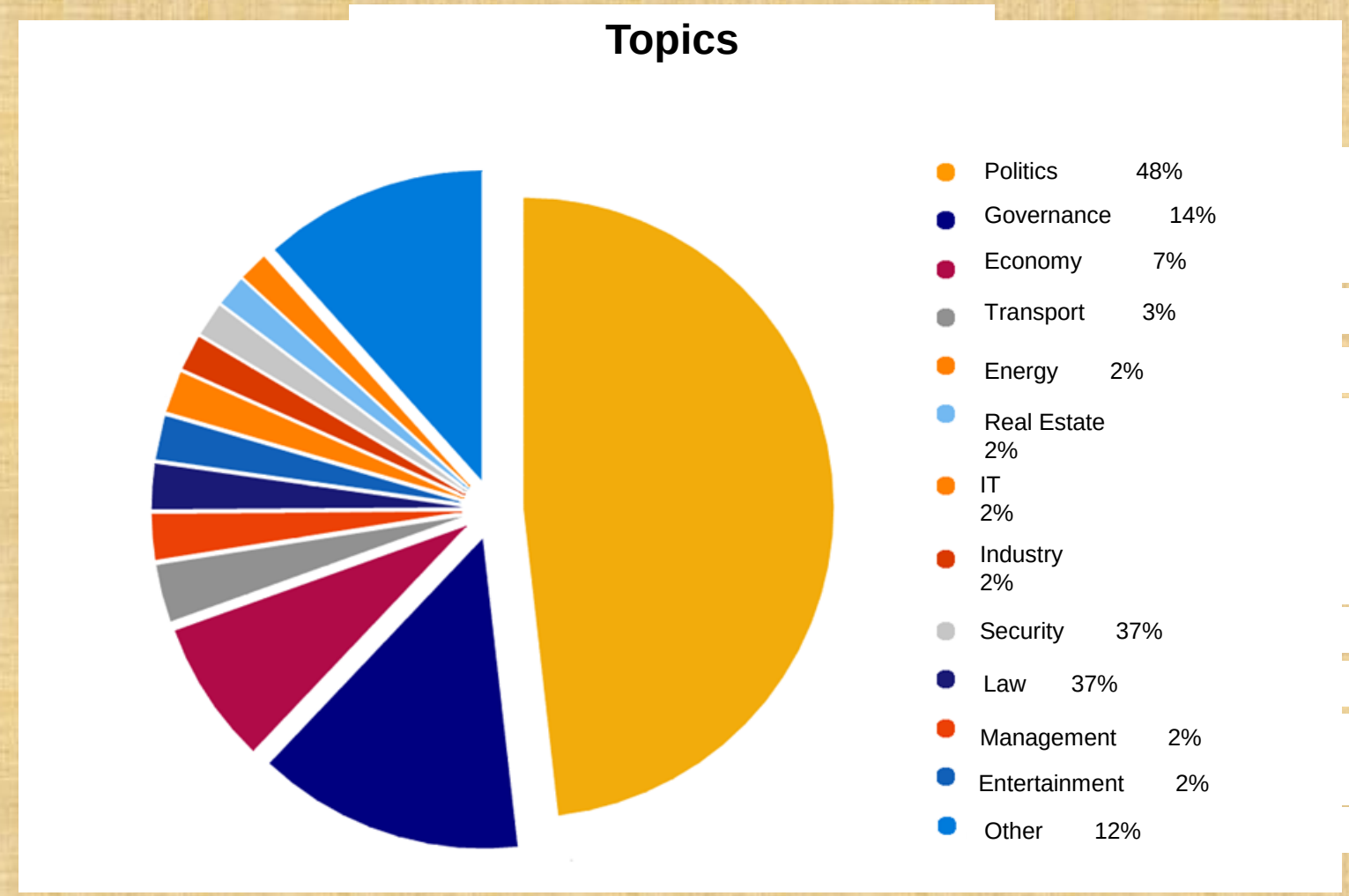
The structure of the selected media - media types



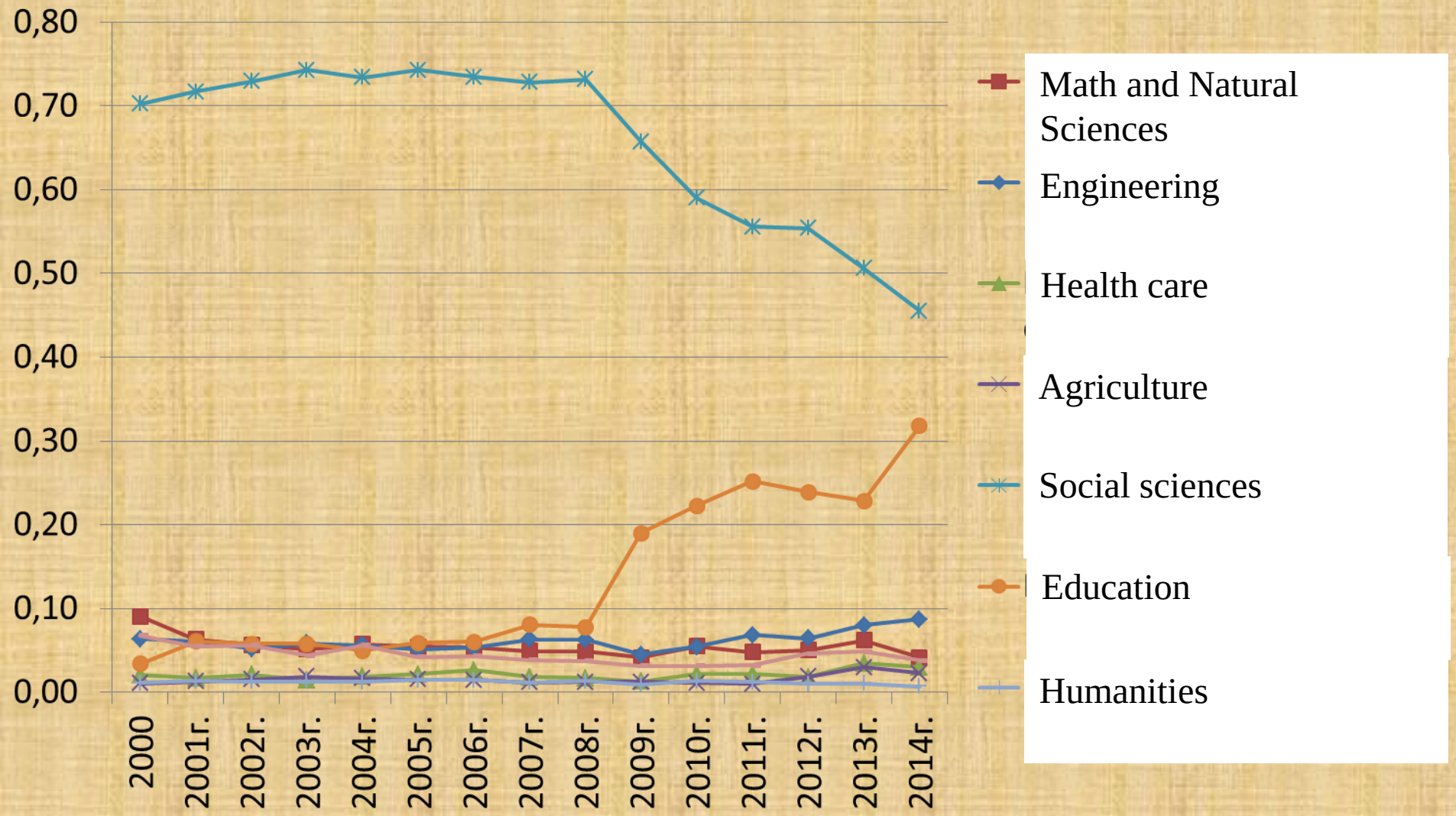
The structure of the selected media - regions



The structure of the selected media - Topics



Part of each profession in total number of citations (in normal scale)



The results of model's calculations

Forecast of the number of applicants

- For modeling choice we must define the number of applicants who exercise choice behavior strategies.
- The basis is a number of graduates from secondary school adjusted with the share of graduates of previous years, which claim the budget places of the current year.
- Only those graduates of previous years who successfully passed the USE in math are taken into account .
- Access to education by the budget remains at 56%.

Forecast of the number of applicants is presented in the table on the next slide.

Forecast of the number of applicants

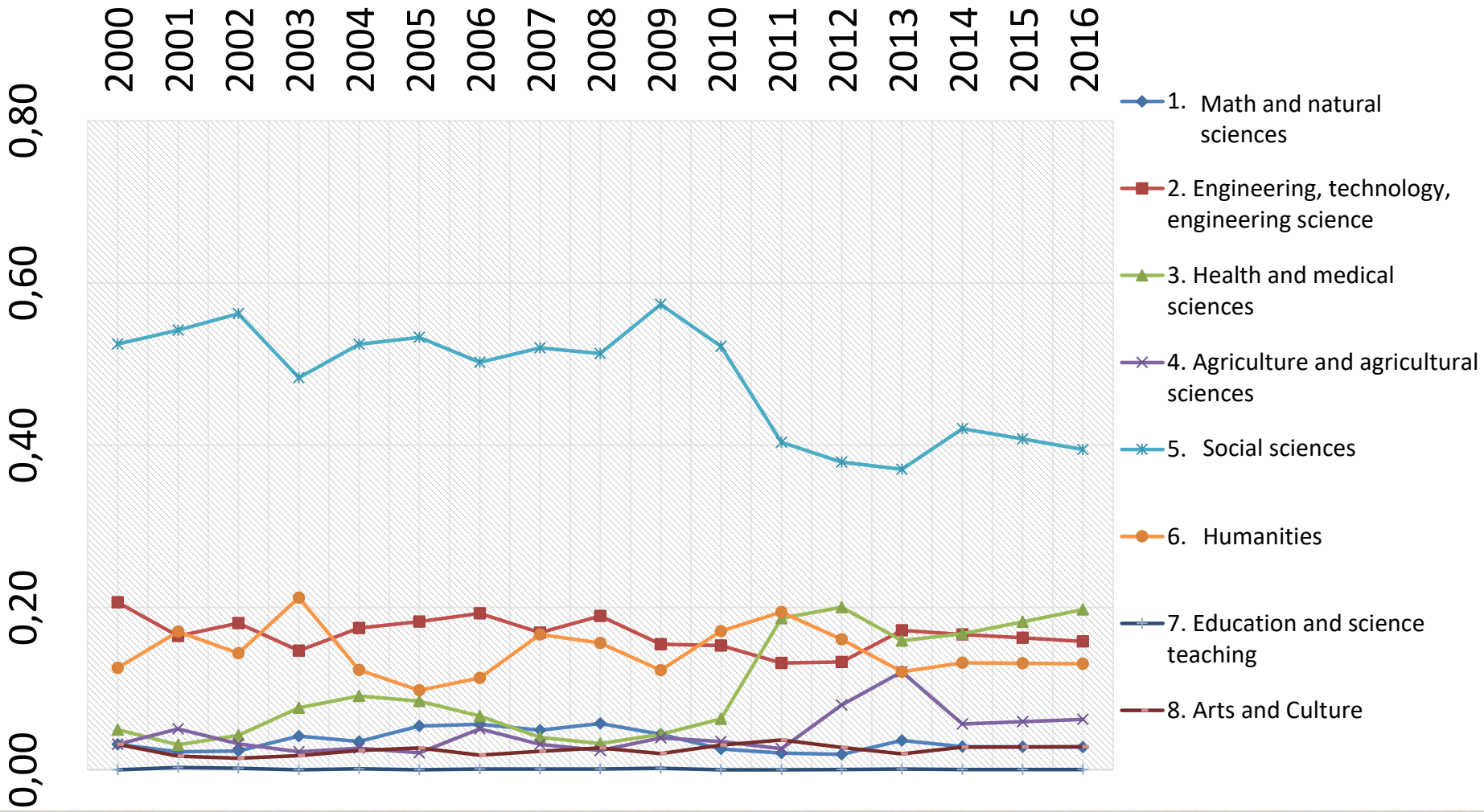
	2011	2012	2013	2014	2015	2016
The number of graduates	654 439	731 745	708 231	659 000	643 184	654 118
Change compared with previous year (%)	0,118	-0,032	-0,070	-0,024	-0,024	0,017
Graduates from previous years, claiming to enroll in bachelor (the share of graduates, %)			0,14	0,18	0,16	0,15
Percentage of graduates from previous years who have passed the USE mathematics (%)			0,48	0,40	0,42	0,45
State guarantee (number of budget places in higher education institutions to number of school graduates ratio), %			0,52	0,56	0,56	0,56
Number of budget places in higher education institutions guaranteed by the state			393 029	394 873	384 387	391 032

Step 1: Select the specialization by applicants - the probability calculated by the model

Specialization	The probability of selecting specialization in % of those who passed the USE in mathematics and Russian language
«Physics»	57,70%
«Poets»	42,30%

photo

Attention given by the state, corporations and investors to the profession, and government policies and programs
2000-2013: fact; 2014-2016: forecast



University efficiency evaluation, application of DEA and sequential exclusion of alternatives

co-authors

Abankina I., NRU HSE
Belousova V., NRU HSE
Bonch-Osmolovskaya A., NRU HSE
Zinkovsky K., NRU HSE
Ogorodniychuk. D, NRU HSE
Petrushchenko V., NRU HSE
Yakuba V., NRU HSE, ICS RAS

The most general idea of what efficiency is

$$Efficieny = \frac{\text{INPUT}}{\text{OUTPUT}} \longrightarrow min,$$

where Input is some aggregated parameter of resources, Output of results. Another form of the same idea

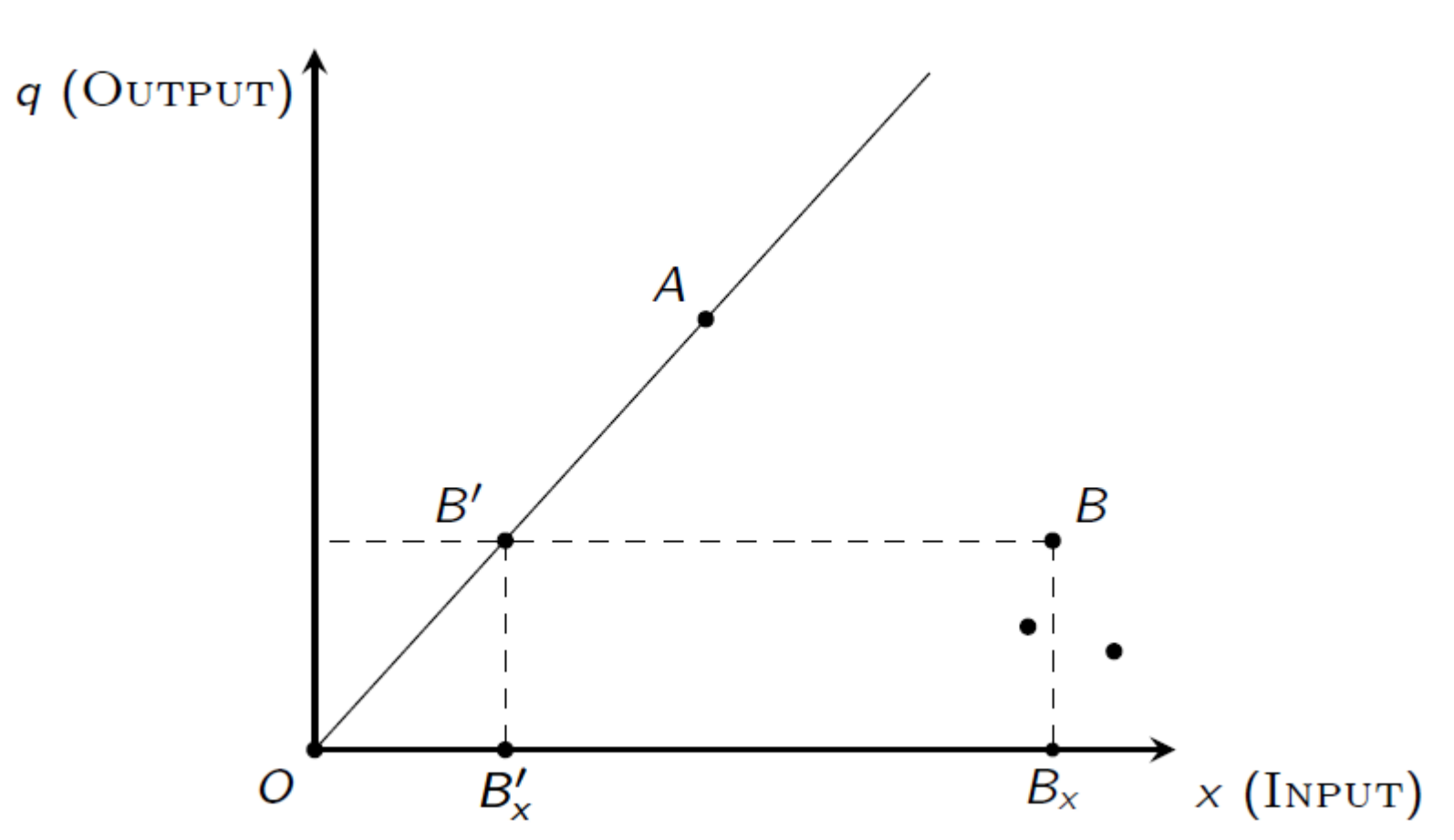
$$Efficieny = \frac{\text{OUTPUT}}{\text{INPUT}} \longrightarrow max.$$

The sample of universities may have several resources and several results.

The key question

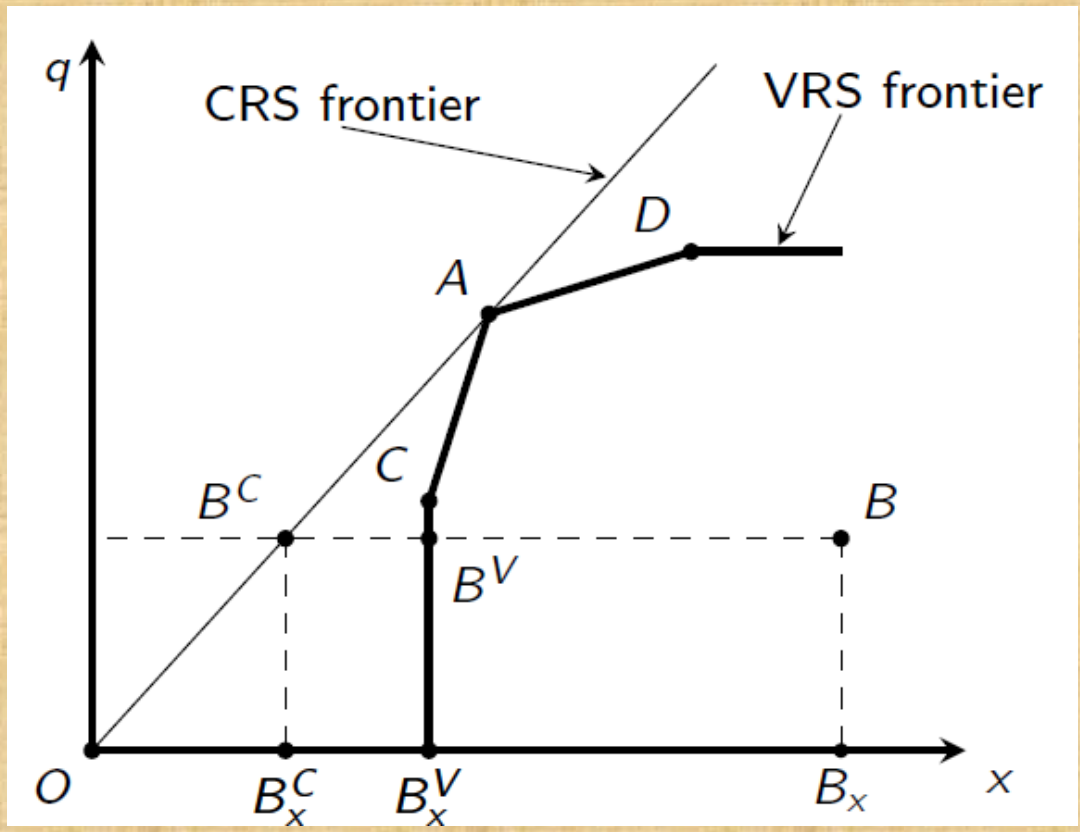
How to define what the Input and Output are in case of several resource parameters and several result variables?

The graphical representation of the simplest case



Efficiency of A is 1. Efficiency of B is the ratio $\frac{|OB'_x|}{|OB_x|}$.

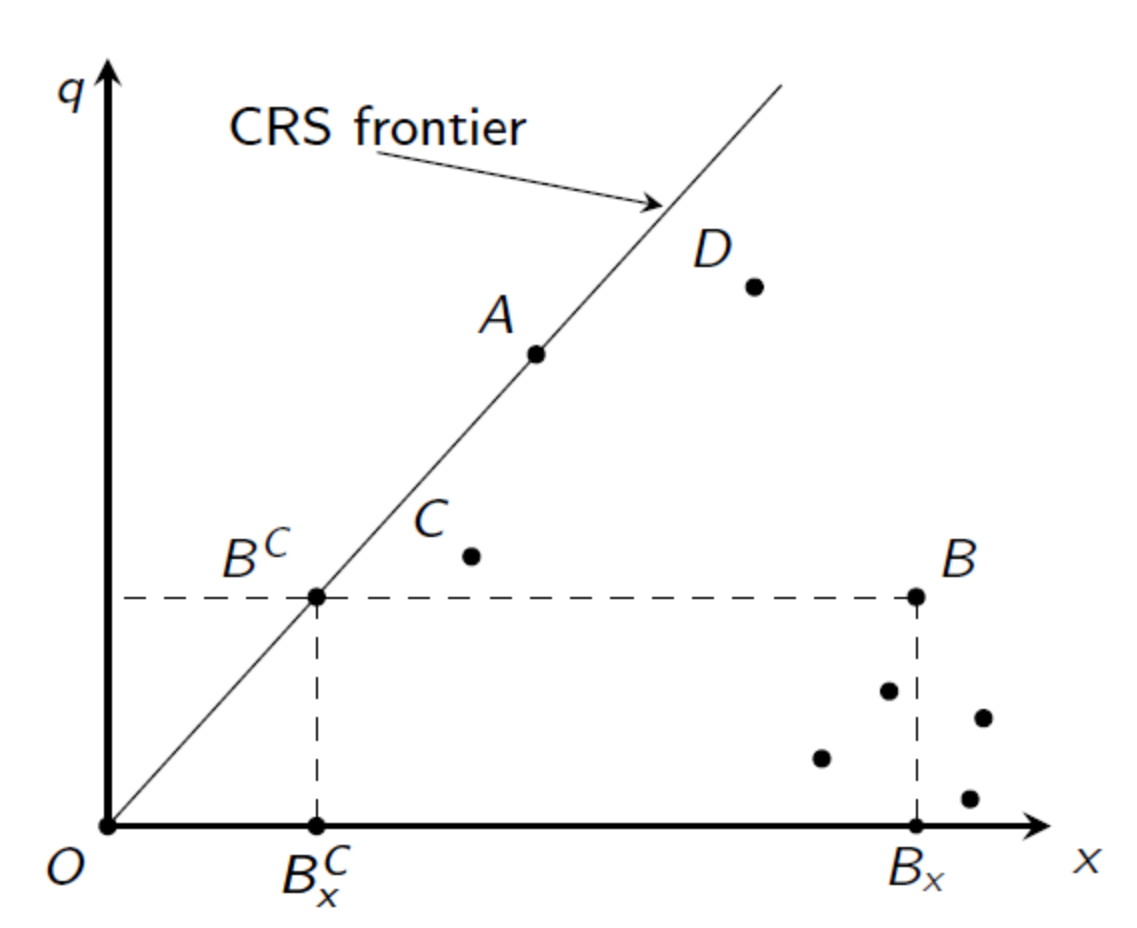
Graphical representation of efficiency frontiers and DEA VRS model.



VRS efficiency of B - $\frac{|OB_x^V|}{|OB_x|}$,

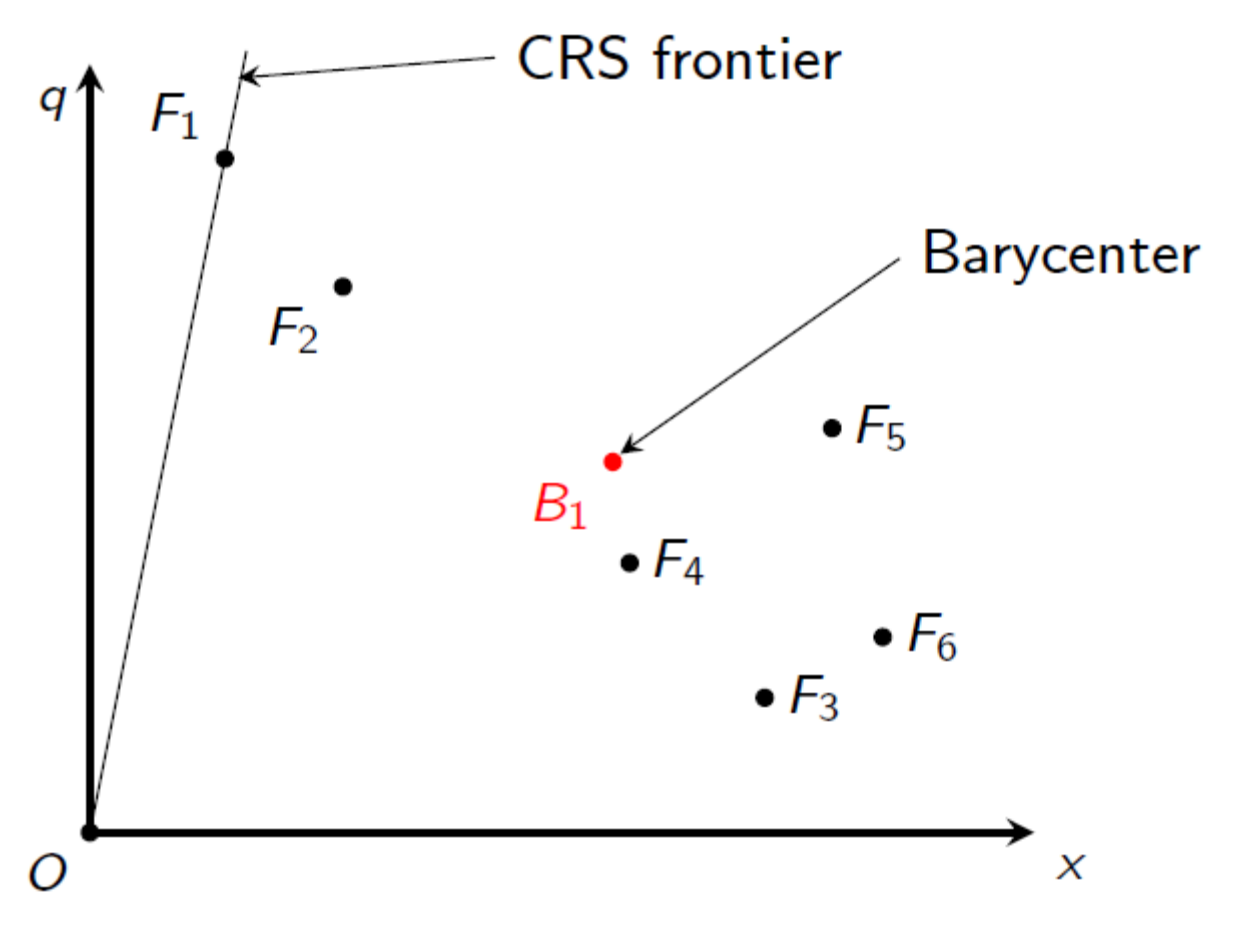
CRS efficiency of B - $\frac{|OB_x^C|}{|OB_x|}$.

What is heterogeneity? How does it appear? Consider the simplest case

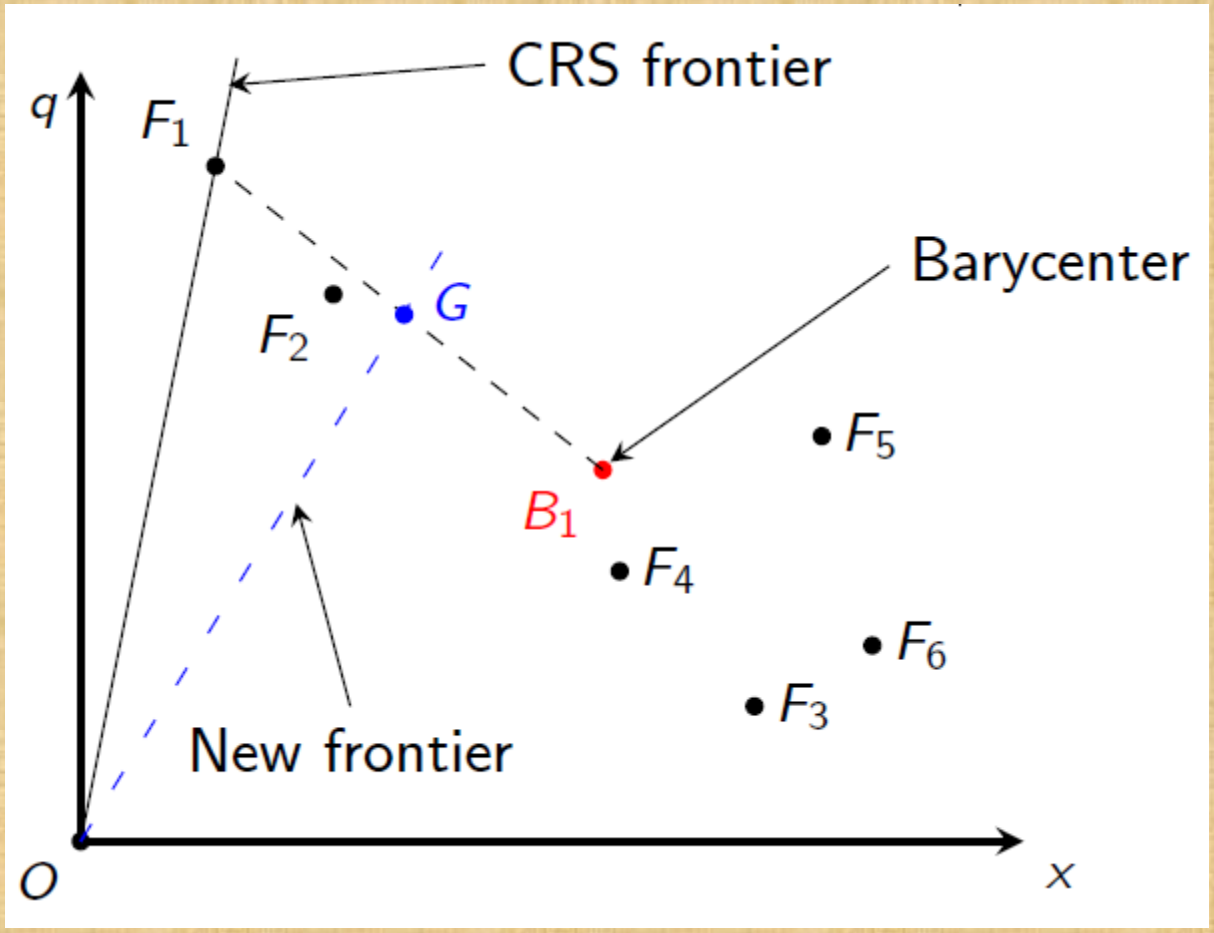


There is a group of universities which are very far from the efficiency frontier.

Graphic representation of the offered algorithm



Graphic representation of the offered algorithm



Algorithm is generalized to the case of arbitrary number of inputs and outputs.

Reputation evaluation

- We have L alternatives and P criteria of reputation evaluation,
- Each alternative is mentioned some number of times in the Internet by each criterion,
- Frequency of mentioning for i-th university regarding j–th criterion presents the number $1 \le a_{ij} \le \omega$ (the higher the better), where ω is an integer.

For every alternative (i = 1, . . . , L) we construct the vector

$$V_i = (v_{1i}, \ldots, v_{\omega i}),$$

where v_{si} the number of criteria, in which i-th alternative obtained s scores (s = 1, . . . , ω).

To compare *i*-th and *j*-th universities we use a threshold procedure

- $u_i \succ u_j$ if $v_{1i} < v_{1j}$, i.e. i -th university obtains less 1-grades than j -th university,
- If $v_{1i} = v_{1j}$ we go to the next components of V_i and V_j , namely v_{2i} и v_{2j} , the conclusion is analogous to the first step,
- We continue the process until on some step we reach an inequality (if there are no inequalities then we consider universities indifferent).

To compare *i*-th and *j*-th universities we use a threshold procedure

- $u_i \succ u_j$ if $v_{1i} < v_{1j}$, i.e. i-th university obtains less 1-grades than j-th university,
- If $v_{1i} = v_{1j}$ we go to the next components of V_i and V_j , namely v_{2i} и v_{2j} , the conclusion is analogous to the first step,
- We continue the process until on some step we reach an inequality (if there are no inequalities then we consider universities indifferent).

In our work the number of criteria (P) equals 13
and
the number of grades (ω) equal 4.

Example. Suppose we have 2 alternatives, 3 criteria and each criterion is evaluated between 1 and 3.

Alternative	1-st criterion	2-nd criterion	3-rd criterion
First university	1	2	1
Second university	1	3	1

Clearly, $V_1 = (2, 1, 0)$ and $V_2 = (2, 0, 1)$.

The first components of two vectors are equal, the second component in V_2 is less than in V_1, the conclusion: the first university is worse than second one.

For application of DEA we use three inputs and two outputs (the data is for pilot sample of 29 universities).

Inputs

The ratio of budget financing to the number of state-financed students,

The ratio of Doctor and Candidate of Science in the teaching stuff,

The quality of entrants (mean value of USE – mandatory exam for all entrants).

For application of DEA we use three inputs and two outputs (the data is for pilot sample of 29 universities).

Inputs

The ratio of budget financing to the number of state-financed students,

The ratio of Doctor and Candidate of Science in the teaching stuff,

The quality of entrants (mean value of USE – mandatory exam for all entrants).

Outputs

Ratio of non-budget income to the number of students who pay tuition,

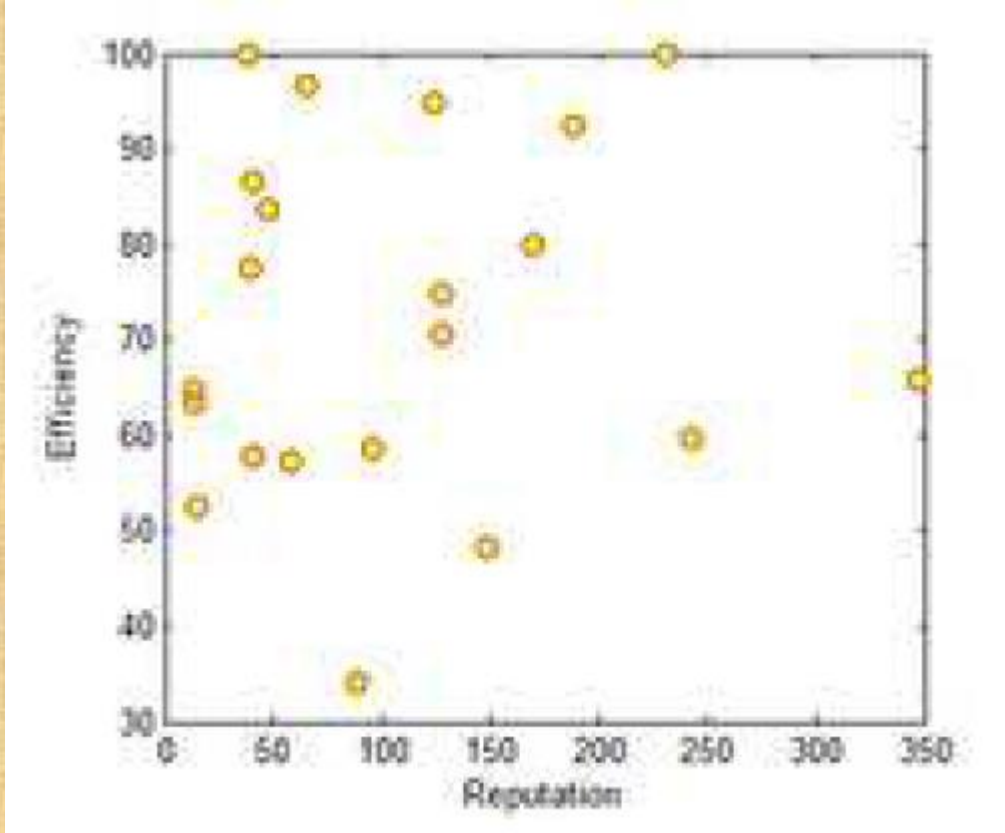
Rating of scientific and publication activity (published by HSE).

Data is taken for 2008-09 year.

13 criteria

- Education,
- Sport, culture, social activity,
- State affiliation,
- Employment,
- Business,
- Science and innovations,
- Religion,
- Finance,
- University infrastructure,
- Expertise,
- Students,
- Alumni,
- Scandals (the higher the worse).

The scatter-plot of efficiency and reputation scores



The right bottom is empty, it may testify on existence of the .efficiency-reputation. frontier. However, this conjecture has to be carefully checked.

Thank you!

20, Myasnitskaya str., Moscow, Russia, 101000
Tel.: +7 (495) 628-8829, Fax: +7 (495) 628-7931
www.hse.ru