Efficiency of R&D spending at the US Universities.

Arina Bogomolova¹,², Igor Balk, Evgeniy Semenov, Natalya Ivashchenko, Anatoly Temkin

¹Lomonosov Moscow State University, GSP- 1, Leninskie Gory, Moscow, 119991, Russia
²Global innovation Labs LLC, 258 Harvard Str #352 Brookline MA 02446 USA
³Irkutsk National Technical Research University, 83 Lermontov Str 664074 Irkutsk Russia
⁴Boston University Metropolitan College, 808 Commonwealth Avenue, Boston, MA, 02215, USA

e-mail: info@innovationlabs.net

Abstract. This article focuses on the efficiency of using university R&D funding. As a parameter of efficiency, we took the so-called “article cost” - the amount of money received for research in the previous year to get 1 article in the current year. It turned out that the average cost of an article at an American university is $101 thousand over the past 8 years. After analyzing a set of universities, it became obvious that the article cost is stable for each university over time, respectively, there are some non-random parameters that affect this value. We have also identified several parameters that, logically, can influence the university performance indicator. They are coefficient of specialization, number of students, number of staff and student per staff ratio, number of articles and others. We also examined several other parameters and made a statistical review of a typical U.S. university with a large research budget. At the end of the article, there is a least squares model that, using these parameters, explains more than half of the differentiation of the efficiency parameter.

1. Introduction

Historically, the main source of funding for higher education is the state, which used to distribute resources depending on the number of students at the university. Now the situation is changing, especially regarding scientific activity. Universities began to sponsor research projects by themselves and attract industry investors. University R&D funding increased across all funding sources in FY 2017. Total R&D expenditures reached $75.3 billion [1].

Not only a lot of money is involved in this area, but also a lot of people. In US there were 6.9 million scientists and engineers in 2016 [2]. Many students at universities are also involved in research activities. Thus, the topic of scientific research is large-scale and relevant in social and economic terms. At the same time, this is a complex and controversial topic, since scientific research often has unpredictable results, which makes it difficult to clearly develop assessment parameters.
Thus, there is a need to create a universal university R&D evaluation method. University can compete for funding and talented students, like any other organization is competing for resources, if it is effective. However, in such a complex system of scientific projects and unique features of each university, there is no definite answer on how to measure the effectiveness of the use of money and labor in university R&D.

There are many articles on the university science funding, but most often they speak of absolute numbers or growth rates [1], but few articles are devoted to the efficiency of the financial resources allocation. There is an example of such a study of university R&D in China, where authors used a network data envelopment analysis [3]. But we will try to go the other way. We will use a universal way (qualitative research articles) to evaluate research results on a sufficient sample of diverse US universities. Earlier in our work [4], we used the optimization model to study the dynamics of universities R&D funding from various sources. Now we turned to the efficiency of using funding and the parameters that affect it.

2. Methodology and data

There are many features that affect research results. The most interesting topic is the choice of the main efficiency indicator for R&D activity in university. Of course, we cannot trace the use of money in each project at all universities, so we will take aggregate indicators to test the hypothesis.

First, we need to know each university R&D funding. The primary source of information on R&D expenditures at U.S. colleges and universities is collected by National Center for Science and Engineering Statistics [5]. For our study, we took data on the 53 universities with the largest R&D funding for the period 2008-2017.

We want to measure scientifically significant results, which are usually published in international article databases, such as Scopus. Constant peer review of the journals included in the Scopus database allows us to ensure high quality standards of articles. The data on the Articles for the period 2008-2018 we took directly from Scopus on each university studied [6].

The efficiency indicator is calculated as follows: funding allocated to R&D at the university divided by the number of articles. But we assume a time lag, since it takes time to use the money received for research. Moreover, the publication of the article usually takes about six months. Thus, we believe that efficiency in 2018 is caused by money allocated in the previous year. The value that we received can be interpreted as the “article cost” – Cost – basically the amount of money resieved in the previous year that is needed to get an article current year. All articles are taken from the database of the same high quality, respectively, the lower the cost of the article, the more efficiently the funds are used. Therefore, we choose this quantitative and universal indicator to evaluate finance usage efficiency.

There is another important point, studies in different areas require a different amount of time and labor. For example, research in molecular biology requires more time and equipment than a statistical sociological survey. However, since we have taken many universities, and each of them is engaged in many studies in different fields, so the average value will be objective and can be compared between
universities. Nevertheless, we will consider such parameter as the specification of the university’s scientific activity. Using Scopus data, we calculated specification coefficient:

- **Specification.** Share of articles in the most popular specialty of the university + 0.5 * Share of articles in the top 2 of the popular specialties.

The next group of variables should describe the overall performance of the university. We used data gathered by Times Higher Education (THE) [7] and College Factual [8] about the following features:

- **Students.** We suggested that the number of articles may depend on the size of the university. Therefore, we took such a parameter as the number of full-time students.

- **Students per Staff.** University staff have two main tasks - student education and research (and as a result, the publication of articles). Thus, the lower the students per staff ratio, the more time professors have for research.

Using previous variables, we can find the number of university staff (Staff) by dividing number of students by students per staff ratio.

We also have data for each university for the following parameters: Out-of-state Tuition and Fees, Percentage of International Students and Percentage of Females students. This data was also taken from THE university rating. We will also be interested to see the number of articles per 1 teacher and 1 student at each university. Since we took the 53 most funded universities, it will be interesting to look at these parameters’ averages, which let us see the big picture.

So, let's move on to the methods of analyzing the collected information. First, we use the statistical method to find average values and extrema. Then, we analyze the consistent pattern between our variables using Pearson correlation coefficient [9], which can be described by formula:

\[ r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \]

where n is the number of samples, \( x_i, y_i \) are the samples with respect to the index i, \( \bar{x} \) and \( \bar{y} \) are mean values of the sample.

Finally, we will use the least squares method [10] to evaluate the causal relationship between the dependent variable (Cost) and a group of independent variables, as well as their joint explanatory ability. Least squares method is minimization of the following function:

\[ \sum_{i=1}^{n}(y_i - f_i(x))^2 \rightarrow \min \]

A simple data set consists of n points (data pairs) \( x \) is a set of independent variables and \( y_i \) is a dependent variable whose value is found by observation. \( f_i(x) \), \( i = 1, ..., m \), \( m > n \) is a set of functions of x. The task is to select such parameters of \( f_i(x) \) so that the values of these functions are as close as possible to \( y_i \). So, the aim is to check whether the selected parameters affect the article cost individually or jointly.
3. Practical study

Let's move on to statistics. We found that the average article cost over the last 8 years is $101 thousand, $105 thousand in 2018. The dynamics of article cost by university (sorted by article cost in 2018) seems stable (Fig.1) - the spread over the years is small. The darker the color of the dot, the newer the value of the variable. So, this is obvious, there should be factor(s) to explain such a diversification among universities and why Harvard spend only $43 thousand to get an article next year, but for example University California needs almost 4 times more.

Figure 1. Article cost by university for 2011 – 2018.

Let’s introduce a typical university with high R&D funding. Among the universities studied average Specification is 0.25, so almost every fourth article is written on one of the 2 most popular topics at the university. The most popular well-funded specialties are medicine, biochemistry, genetics, molecular biology, engineering, physics and astronomy.

We did not manage to find comprehensive information on the following parameters of some universities, so we ruled out incomplete observations. Thus, 46 universities remain in study. There are on average 32 thousand students, of which 17% are foreigners and 49% are women, and about 3 thousand staff (so that's about 11 students per staff). In our sample, there are both small universities for 11 thousand students (MIT), and huge ones for 65 thousand students (Indiana University). There are 3 articles per staff and 0.3 per student in 2018. Out-of-state Tuition and Fees is $38 thousand on average (from $17 thousand in University of South Florida to $52 thousand in University of Southern California).
To study the correlation of the factors we have chosen, let us look at the correlation matrix (Fig. 2). It presents the pairwise Pearson correlation coefficients.

|                  | Cost 2018 | Specification | Articles in 2017 | Students | Staff | Students per Staff | Tuition and Fees | International students | Female students |
|------------------|-----------|---------------|------------------|----------|-------|--------------------|-------------------|------------------------|-----------------|
| Cost 2018        | 1,00      | 0,45          | -0,62            | -0,19    | 0,02  | -0,10              | 0,00              | -0,04                  | 0,00            |
| Specification    | 1,00      |               | -0,29            | -0,48    | 0,07  | -0,03              | -0,36             | 0,05                   | 0,19            |
| Articles in 2017 |           | 1,00          |                  | 0,13     | 0,00  | 0,00               | 0,05              | 0,19                   | 0,02            |
| Students         | 1,00      |               |                  | 0,46     | 0,58  | -0,49              | -0,28             | 0,21                   |                 |
| Staff            | 1,00      |               |                  | -0,39    | 0,01  | 0,00               | -0,14             | 0,39                   |                 |
| Students per Staff| 1,00    |               |                  | -0,57    | 0,63  | -0,18              | -0,11             |                       |                 |
| Tuition and Fees |           |               |                  | 1,00     | 0,63  | -0,18              | -0,11             |                       |                 |
| International Students | 1,00    |               |                  | 0,00     |       |                    |                   |                       |                 |
| Female students  |           |               |                  | 0,00     |       |                    |                   |                       | 1,00            |

**Figure 2.** Correlation matrix

There is more or less significant correlation between the article cost and such factors as specialization and the number of articles. Some correlation exists between the number of students and the article cost, possibly because students are workforce for writing them. The remaining parameters practically do not correlate with our target variable. It is rather unexpected to see that the correlation is so small. It turns out that the other variables do not affect the article cost. But we can check by the least squares method how many percent of differentiation we can explain with the chosen parameters. Another point worth noting is the high correlation coefficient between the number of students, staff and student per staff ratio, so we will not use them in one model.

The application of the least squares method to statistical data produced the following results (Fig. 3).

|                  | Model 1       | Model 2       | Model 3       | Model 4       |
|------------------|---------------|---------------|---------------|---------------|
| const            | 105***        | 101***        | 110***        | 91***         |
|                  | (9,87)        | (15,13)       | (12,74)       | (14,07)       |
| Specification    | 91***         | 97,91***      | 85,99***      | 108,49***     |
|                  | (20,50)       | (27,59)       | (22,08)       | (23,78)       |
| Articles in 2017 | -0,003***     | -0,01***      | -0,003***     | -0,003***     |
|                  | (0,00)        | (0,00)        | (0,00)        | (0,00)        |
| Students         | 0,001         | 0,001         | 0,001         | 0,001         |
|                  | (0,00)        | (0,00)        | (0,00)        | (0,00)        |
| Staff            |               | -0,001        |               | 0,71          |
|                  |               | (0,002)       |               | (0,51)        |
| Students per Staff|             |               |               |               |
|                  |               |               |               |               |
| R² adj           | 0,55          | 0,54          | 0,54          | 0,55          |

**Figure 3.** Comparison of Least Squares Models.

*** - coefficient is significant (where α=0,01).
The coefficient of determination shows that we can explain more than half of the vibrations of the dependent variable. However, the addition of variables describing the number of students, staff and their ratio did not lead to improvements. Moreover, all these variables were not significant. Thus, we can opt for Model 1.

4. Conclusion

The R&D funding efficiency can be estimated by such a variable as the article cost. Having studied this variable, we concluded that it is quite stable for each university and set a goal to find out what factors influence this performance indicator. We checked several factors that depict university that could influence this variable. We found that the cost per article is not affected by the number of students, staff, and student per staff ratio, which reflects the professor’s worktime for research. Moreover, we were able to demonstrate more than half of the differences in the article cost by using the number of articles in the previous year and the specialization coefficient.

The next step in our study is the selection of other factors that may affect the R&D funding efficiency and the following modification of the model 1.

References

[1] Gibbons M. T. 2018 R&D Expenditures, by Source of Funding National Science Foundation URL: https://www.nsf.gov/statistics/2019/nsf19302/overview.htm

[2] John F. Sargent Jr. 2017 The U.S. Science and Engineering Workforce: Recent, Current, and Projected Employment, Wages, and Unemployment Congressional Research Service vol 7-5700

[3] Xionghe Qin, Debin Du 2018 Measuring universities’ R&D performance in China’s provinces: a multistage efficiency and effectiveness perspective Technology Analysis & Strategic Management, 30:12, 1392-1408, DOI: 10.1080/09537325.2018.1473849

[4] Bogomolova A, Balk I, Ivachenko N 2017 Budget optimization modeling for sustainable development of the university research: the example of USA IOP Conference Series: Earth and Environmental Science Vol.73

[5] Higher Education Research and Development Survey Fiscal Year 2017 National Center for Science and Engineering Statistics URL: https://ncsesdata.nsf.gov/herd/2017/index.html

[6] Scopus URL:https://www.scopus.com

[7] Times Higher Education World University ranking URL: https://www.timeshighereducation.com

[8] College Factual, URL: https://www.collegefactual.com

[9] Rodgers J. L. Nicewander W. A. 1988 Thirteen ways to look at the correlation coefficient The American Statistician 42(1) pp 59-66
[10] Bretscher O 1995 Linear Algebra With Applications (3rd ed.). *Upper Saddle River, NJ: Prentice Hall*