Journal of Informetrics 9 (2015) 183–196

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Distributions of scientific funding across universities and research disciplines

Jiang Wu [a,b,*]

[a] School of Information Management, Wuhan University, Wuhan, Hubei 430072, China
[b] Institute of Automation, Chinese Academy of Sciences, Beijing 100081, China

ARTICLE INFO

Article history:
Received 24 August 2014
Received in revised form
20 December 2014
Accepted 23 December 2014
Available online 14 January 2015

Keywords:
Gini coefficient
NSFC
Distribution of scientific funding
Distribution fits
Inequality of funding allocation

ABSTRACT

Distributing scientific funding to the suitable universities and research fields is very important to the innovation acceleration in science and technology. Using a longitudinal panel dataset of the National Natural Science Foundation of China (NSFC), the total 224,087 sponsored projects is utilized to investigate the distributions of scientific funding across universities and research disciplines. The inequality of funding distribution is studied through the investigation of Gini coefficient, and its fundamental rules are discovered through the technique of distribution fitting. It is found that the inequality of distributions of NSFC funding across 1971 universities is decreasing, and the distribution of funding and supported universities of 971 research fields follow Generalized Pareto distribution and Geometric distribution function, respectively. This study is dedicated to give an entire landscape to help make policy of distributing scientific funding.

1. Introduction

Scientific funding plays a key role in individual scientific research, university discipline construction and national innovation system (Lok, 2010). China is rising as a major contributor to science and technology in the world (Xie, Zhang, & Lai, 2014). National Natural Science Foundation of China (NSFC), as the main financial support for natural scientific research, has invested more than 100 billons into more than 300 thousands projects to support about 1 million researchers, since it established in 1986. The biggest challenge of the scientific foundation such as NSFC is how to efficiently and fairly allocate such a large amount of funds to different individuals, disciplines and universities (Lane, 2009; Pendlebury, 1991). In this paper, the distributions of scientific funding will be investigated, and we would like to find the fundamental laws of these distributions.

Previous research of scientific funding usually focus on peer review of the funding (Roebber & Schultz, 2011) and the performance evaluation of sponsored projects (Auranen & Nieminen, 2010). Besides, the inequality of funding allocation has attracted some interests as well in recent years (Halffman & Leydesdorff, 2010; Shibayama, 2011). The academic funding has become increasingly preoccupied with social and economic inequality in many countries (Xie, 2014). Nevertheless, it has been shown that universities have not become more unequal in terms of publications (Halffman & Leydesdorff, 2010), but they become unequal in terms of funding (Shibayama, 2011). The distribution of funding has been studied based on Japanese and USA funding system (Shibayama, 2011; Wu, 2013). To the best of our knowledge, there has not been such a

* Correspondence to: School of Information Management, Wuhan University, Wuhan, Hubei 430072, China. Tel.: +8602768755647.
E-mail address: Jiangw@whu.edu.cn

http://dx.doi.org/10.1016/j.joi.2014.12.007
1751-1577/

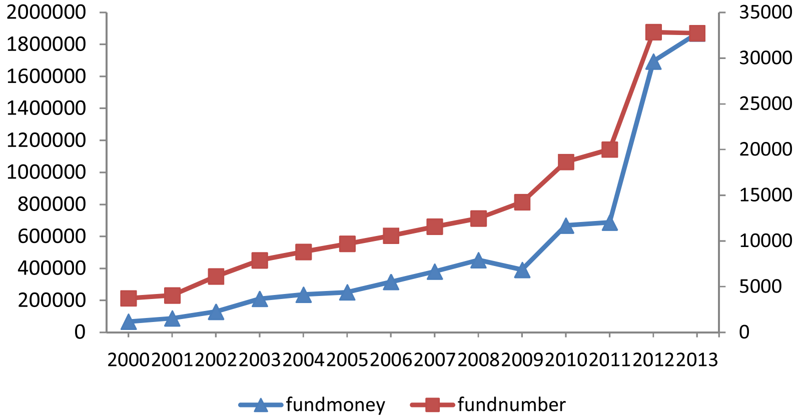

Fig. 1. The growth of scientific funding in terms of the total number of grants (*fundnumber*) and total amount of sponsored money (*fundmoney*).

research that investigates the inequality of NSFC funding, therefore, this paper tries to fill this gap through investigating the distributions of NSFC funding across universities and scientific disciplines.

Furthermore, rather than using the Web of Science bibliometrics data (Xie et al., 2014) and focusing on the analysis at the individual level (Abramo, Cicero, & D'Angelo, 2013), in this paper, the R&D system in China is investigated at the university and discipline level through studying the panel data from the main research funding information system. The performance achieved by the support of funding is important to evaluate the funding allocation policy, and usually the higher productivity of papers and patents means the higher performance in the evaluation (Payne & Siow, 2003). Moreover, it is also interesting to find the intrinsic fundamental rule of the distributions of funding, discover the reason of inequality and know the entire funding landscape of universities and disciplines for better policy making.

Therefore, the main contributions of this paper have been twofold, based on the dataset of sponsored project of National Natural Science Foundation of China. *First*, the longitudinal investigation is dedicated to discover the evolution of distributions of scientific funding. The trend of inequality of funding allocation is paid more attention to be benefit to make a suitable policy. The main finding is that the inequality of distributions of scientific funding across 1971 universities is decreasing from 2000 to 2013 as a whole in terms of both the amount of sponsored money and the number of grants. *Second*, we further investigate the fits of these distributions to find the fundamental law in terms of funding distributions. The main finding is that the distribution of funding of 971 research fields follow Generalized Pareto distribution function and the distribution of supported universities of 971 research fields follows Geometric distribution function.

The rest of this paper is organized as follows. Section 2 presents our methods including dataset collection and metrics calculation. In Section 3, we empirically analyze and visualize the results. Finally, Section 4 summarizes our work, and addresses limitations and possible extensions in the future.

2. Dataset and methods

In order to understand the current allocation of scientific funding, a variety of distributions across universities and scientific disciplines calculated in terms of the total amount of sponsored money and the total number of sponsored projects is investigated. In the following, the dataset and methods are introduced, respectively.

2.1. Dataset

This paper retrievals the raw dataset of all the sponsored projects from 2000 to 2013, taken from the ISIS system (Information System of National Natural Science Foundation of China). The total 224,087 records of sponsored projects are stored into the MySql database for the convenience of analysis. Each record includes the Title, Principal investigator, Approval year, Institution, Amount of sponsored money, Approval number, and Discipline codes, etc.

In the NSFC, there are eight departments: Mathematical and physical science, Chemical sciences, Life sciences, Earth sciences, Engineering and materials science, Information science, Management science and Medical sciences. For each project, applicants need to provide the discipline code, which is used to select the suitable reviewers during the peer review of proposal and classify them. The discipline code is a three-level code to indicate the detailed discipline a proposal belongs to. The total 86 number of first-level discipline codes indicates the research area such as Mathematics, Physics, Mechanics, and Astronautics. The total 981 number of second-level discipline codes indicates the research field such as Algebra, Functional analysis, and Geometry. The total 1679 number of third-level discipline codes indicates the detailed research direction such as Analytic number theory, Algebraic number theory, and Number theory application.

As shown in Fig. 1, the number of grants (*fundnumber*) increases over year and the total amount of sponsored money (*fundmoney*) also increases, with the only exception of the decrease in 2009. To be noticed, *fundnumber* is a representative of

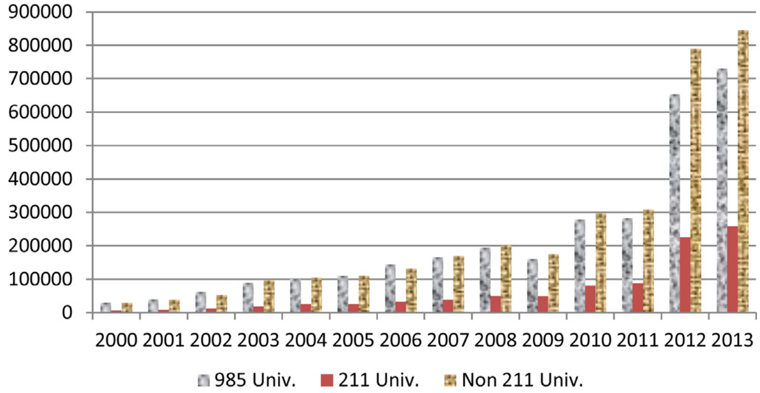

Fig. 2. The growth of scientific funding in 985, 211 and non-211 universities.

the broadness of funding, while *fundmoney* concentrate more on the capacity of funding. They measure different two relative aspects of scientific funding. In Fig. 1, especially, in 2012, more than RMB 17 billons Yuan have been invested, leading to a substantial growth in term of both the number of grants and the total amount of sponsored money. It is mainly because, to accelerate innovation, in 2012 the sponsored money per project increased to an average of 800,000 Yuan for four years, and NSFC set up Excellent Youth Science Fund.

In order to check the distributions of funding across the top-class universities, a dataset to indicate 985 and 211 universities has also been used to distinguish the sponsored projects. Project 211 is the Chinese government's new endeavor aimed at strengthening about 100 institutions of higher education and key disciplinary areas as a national priority for the 21st century (Choi, 2010). Project 985 is a constructive project for founding world-class universities in the 21st century conducted by the government of the People's Republic of China (Zhang, Patton, & Kenney, 2013). The universities involved into Project 211 and Project 985 belong to top-class universities, which will obtain more sponsored money than other universities from the central government. Now the total 39 universities have been selected into Project 985 and the total 112 universities have been sponsored by Project 211. A 985 university is also the 211 university, vice versa not. 985 universities usually rank higher than 211 universities in terms of scientific funding, and are regarded as global-class universities in China (Zhang et al., 2013).

As shown in Fig. 2, the growth of scientific funding in 985, 211 and non-211 universities are investigated. As indicated in our dataset, from 2000 to 2013, the total 1859 non-211 universities have won the sponsorship of National Natural Science Foundation of China. It is shown that the amount of sponsored money obtained by 39 985 universities is up to 1/3, but 73 only 211 universities takes less than 1/6 of the total amount of sponsored money. It is clear that 985 universities are more successful in winning scientific funding, and inequality of distributions exists. In the following, the degree of inequality and its changes are quantitatively investigated further.

3. Methods

In order to catch the inequality existed in distributions of scientific funding, Gini coefficient is taken for the measure of inequality (Halffman & Leydesdorff, 2010). The Gini coefficient is originally used in characterizing income and wealth distributions, and its calculation depends on the Lorenz curve, which shows the percentage of the total wealth given to the bottom *x*% of entities (Dorfman, 1979). The Gini coefficient is a measure of the deviation of the Lorenz curve from the equidistribution line which is a line connecting [0, 0] and [1, 1]. Several previous works have used Gini coefficient in the academia as a measurement of inequality (Halffman & Leydesdorff, 2010; Shibayama, 2011; Xie, 2014). In this paper, we used normalized Gini coefficient (Halffman & Leydesdorff, 2010):

$$G_t = \frac{\sum_{i=1}^{n}(2i-n-1)x_i}{(n-1)\sum_{i=1}^{n}x_i} \tag{1}$$

with n being the number of universities in the dataset and xi being the number of grants or the amount of sponsored money of the university with position i in the ranking. The higher Gini coefficient means the higher inequality of distributions of scientific funding.

Furthermore, in order to understand the fundamental rule of the distributions, distribution fitting technology has been used. After automatic selection of a suitable distribution by calculating goodness-of-fit measures, the Maximum Likelihood Estimation method (MLE) is utilized to estimate the parameters. The parameters estimated in terms of a higher goodness-of-fit can describe and predict the probability or to forecast the frequency of occurrence of the magnitude of the scientific funding in a certain time interval. The above methods will be applied in a longitudinal dataset, and the distribution of funding each year will be investigated separately to catch the evolution of these distributions.

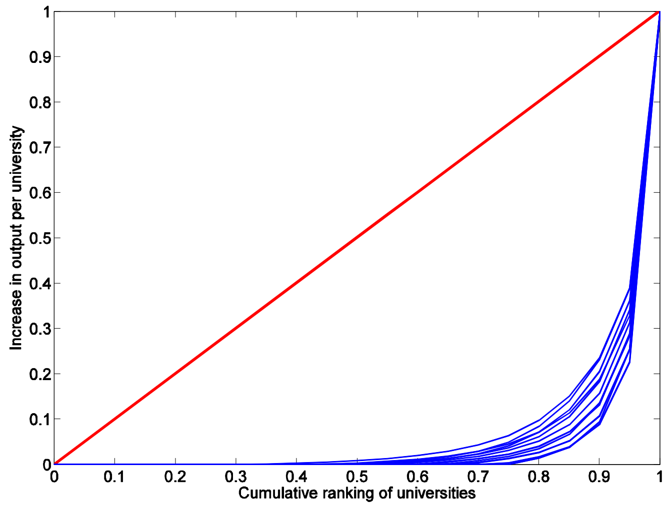

Fig. 3. The calculation of Gini coefficient in term of the amount of sponsored money of universities.

We used Matlab language to do programming of distribution fitting both in discrete and continual format. Essentially the dataset is discrete, but it can be approximately regarded as continual format because of big enough number of samples. Distribution fitting is used to select a statistical distribution that best fits to our dataset, which is presumed to be generated by some random process. We have tried a series of probability distributions such as Generalized Extreme Value, Generalized Pareto, Geometric, Negative Binomial, Poisson, and Power Function. In the distribution fitting, the goodness of fit (GOF) tests shows how well the distribution we selected fits to our data. The procedure consists of defining a test statistic which is some function of our data measuring the distance between the hypothesis and the data, and then calculating the probability of our data which have a still larger value of this test statistic than the value observed, assuming the hypothesis is true (Chatterjee & Hadi, 2012).

Three goodness of fit (GOF) tests, namely Kolmogorov–Smirnov test, Anderson Darling test and Chi-squared test (Chatterjee & Hadi, 2012), are used to compare the GOF of the fitted distributions. The GOF score, namely the probability of test, indicates if it is a good fit or not. Small probabilities indicate a poor fit. Especially high probabilities (close to one) correspond to a fit which is too good to happen very often, and may indicate a mistake in the way the test was applied. In this paper, the distribution with the best score of Kolmogorov–Smirnov test is selected as the proposed fitting distribution. Chi-squared test and Anderson Darling test are used to verify the selection of the best fitted distribution. It is shown the test scores for all the tests in the following are not too large to be over-fitting, which indicates that a complex model with a high GOF score is not a good fit as well. Therefore, through the above three GOF tests, we can make a good trade-off between model fit and model complexity in the distribution fitting.

4. Results

Using the above methods and dataset, in the following, we investigate the distribution of funding across universities, research areas and research fields, respectively. Distributions of supported universities of research fields are investigated as well.

4.1. Distributions of funding of universities

It has been supposed that there exists inequality of scientific funding among the totally 1971 universities, which have won the sponsorship of National Natural Science Foundation of China from 2000 to 2013. Taking advantage of Gini coefficient, we can quantitatively analyze this inequality. As shown in Fig. 3, the red solid line is the equidistribution line, and a series of blue solid curves are Lorenz curves from 2000 to 2013. The fourteen Lorenz curves are close to each other, indicating the minor changes of inequality over the years. The calculation of Gini coefficient depends on the area between Lorenz curve and equidistribution line. The Gini coefficients of each year in terms of the amount of sponsored money and the number of grants are shown in Fig. 4.

As shown in Fig. 4, the Gini coefficient decreases as a whole but still remain a higher level, which indicates a higher inequality. Comparatively, it is reported that the Gini coefficient of federal research funding in U.S. universities increased steadily from 0.75 in 1990 to 0.81 in 2010 (Xie, 2014), which is much less than China. The inequality is considered undesirable (Halffman & Leydesdorff, 2010; Shibayama, 2011; Xie, 2014), therefore the decrease in China's Gini coefficient is a positive development. Especially, in 2012, the Gini coefficient decreases dramatically, however, in 2013, it increases again. It is clear to see in 2009 and 2010 the Gini coefficients in term of the number of grants and the amount of sponsored money has

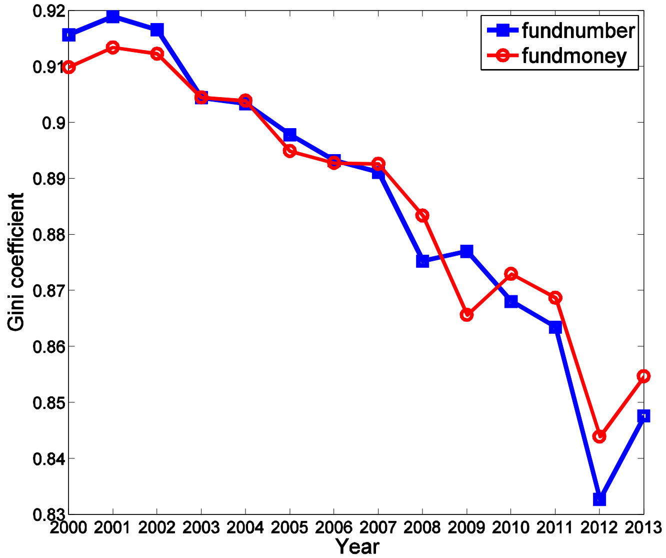

Fig. 4. Gini coefficient in terms of the number of grants (*fundnumber*) of universities and the amount of sponsored money of universities (*fundmoney*).

Table 1
The top 7 universities in terms of both *fundmoney* and *fundnumber*.

1. Shanghai Jiaotong University	5. Fudan University
2. Zhejiang University	6. Huazhong University of Science and Technology
3. Peking University	7. Sun Yat-Sen University
4. Tsinghua University	

Table 2
Research areas and their corresponding codes, area numbers and departments. (Area numbers of 86 research areas correspond to the rows in Figs. 5 and 6.)

Scientific department	Mathematical and physical sciences	Chemical sciences	Life sciences	Earth sciences	Engineering and materials sciences	Information sciences	Management sciences	Medical sciences
Area number	1–5	6–12	13–32	33–38	39–47	48–52	53–55	56–86
Research area	A01–A05	B01–B07	C01–C20	D01–D06	E01–E09	F01–F05	G01–G03	H01–H31

different kinds of trends. Especially, in 2010, Gini coefficient in term of the amount of sponsored money increases but the Gini coefficient in term of the number of grants decreases. It indicates that some universities obtain less money per grant although they won the more number of grants. It is also shown in 2012 the Gini coefficient decreases dramatically because of the substantial growth of the number of grants and the total amount of sponsored money as indicated in Fig. 1.

The 985 universities achieved much better performance than the 211 universities in terms of *fundnumber* and *fundmoney*. Among 39 985 universities, the top seven universities ranked in the list of top ten every year from 2000 to 2013, as shown in Table 1. The amount of sponsored money of the top ten universities is about 19% of the total amount of sponsored money of the total 1971 universities in each year. Therefore, the concentration of the allocation of scientific funding is obvious in China.

4.2. Distributions of funding of research areas

In our dataset, there are the total 86 first-level discipline codes that indicate the research area such as Mathematics, Physics, Mechanics, and Astronautics in eight departments as shown in Table 2. The number of grants of research areas and the amount of sponsored money of research areas across years from 2000 to 2013 have been investigated. In order to show the differences among different research areas across years, we use the heat map to visualize the landscape of distributions of funding. Because there is not enough large number of research areas, it is not suitable to use traditional curve fitting in this samples. However, heat map can be a good graphical representation of data where the individual values contained in a matrix are represented as colors, with the ability of visualizing three-dimensional data in the 2D space.

As shown in Figs. 5 and 6, which are in the format of heat maps, the vertical axis indicates the research areas where the area number represents the corresponding research areas in the eight departments as shown in Table 1, the horizontal axis indicates the year from 2000 to 2013, and the color indicates the ratio of the number of grants (*fundnumber*) and the amount

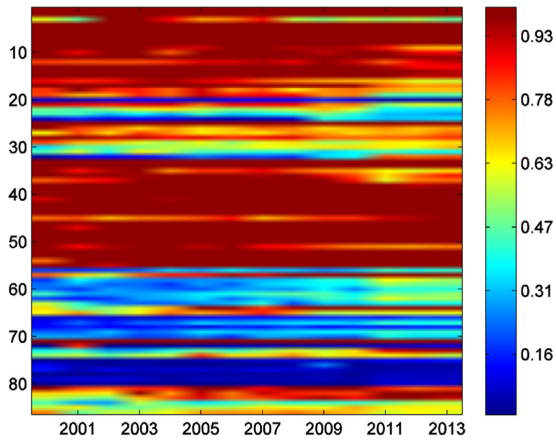

Fig. 5. The ratio of the number of grants (*fundnumber*) of each research area to the total number of grants in all the 86 research areas, denoted by the corresponding area numbers from 1 to 86 as shown in Table 2. (The *y*-axis denotes area numbers, *x*-axis denotes years and the color in each cell denotes the ratio of the number of grants.) (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

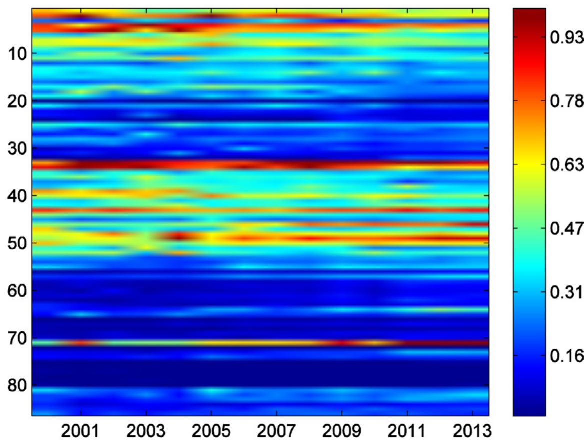

Fig. 6. The ratio of the amount of sponsored money (*fundmoney*) of each research area to the total amount of sponsored money in all the 86 research areas, denoted by the corresponding area numbers from 1 to 86 as shown in Table 2. (The *y*-axis denotes area numbers, *x*-axis denotes years and the color in each cell denotes the ratio of the amount of sponsored money.) (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

of sponsored money (*fundmoney*) of each research area to the total *fundnumber* and *fundmoney* of all the research areas, respectively.

First, as shown in Fig. 5, the color of the area number from 1 to 20 and 33 to 55 is much lighter than other regions. Therefore, the *fundnumber* of the Mathematical and physical science, Chemical sciences, Earth sciences, Engineering and materials science, Information science and Management science is relatively higher than others. There are several exceptions in term of the ratio of the number of grants. The 20: C08 (Immunology) in Life sciences obtains a relatively low *fundnumber* each year. It may be a part of reasons why the world publication count in China is only 16% of USA in the Immunology (Xie et al., 2014). In the Medical sciences from 56 to 86, the *fundnumber* is much lower than other departments. However, the 71: H16 (Oncology) obtains much higher *fundnumber* each year, and the 83: H27 (Traditional Chinese Medicine) and 84: H28 (Science of Chinese Pharmacology) are paid more attentions as well in term of *fundnumber*, mainly because of the national support and protection to the Chinese traditional medicine.

Second, as shown in Fig. 6, the ratio of the amount of sponsored money (*fundmoney*) of each research area indicates a totally different landscape from the landscape of the ratio of the number of grants. The *fundmoney* in the majority of the regions of corresponding research area is low. The two regions with the area number from 1 to 12 and 33 to 55 have relatively higher *fundmoney*. Only a small number of research areas obtain very high *fundmoney*, i.e., 33: D01 (Geography), 34: D02 (Geology) and 71: H16 (Oncology) obtains extremely higher *fundmoney* than others.

Third, the top 10 research areas in term of *fundnumber* and *fundmoney* are also listed in Figs. 7 and 8, respectively. The 86 research areas and their corresponding codes, area numbers are referred to Appendix A. It is shown in the recent years H16 (Oncology) is paid more and more attention because Cancer is one of the most deadly diseases, and the risk of cancer is increasing as well. Moreover, D01 (Geography) ranked top 5 research areas each year and obtains the most amount of

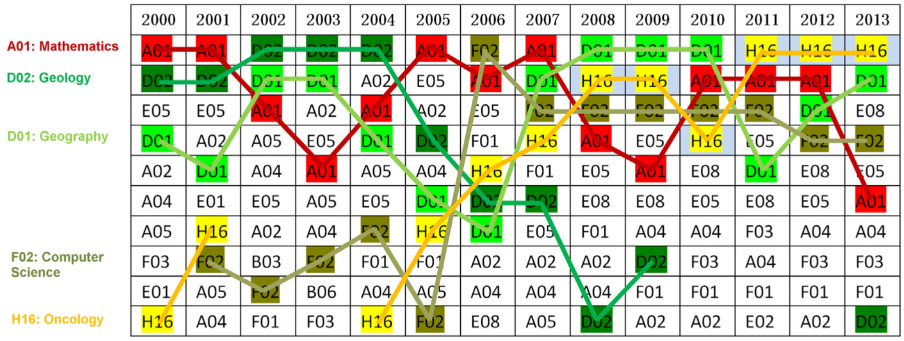

Fig. 7. The top 10 research areas in term of *fundnumber* in each year (the curves represents the changes of ranks of the research areas that ever ranked No. 1 from 2000 to 2013 in term of *fundnumber*). Note: The corresponding names of the other codes of research areas are referred to Appendix A.

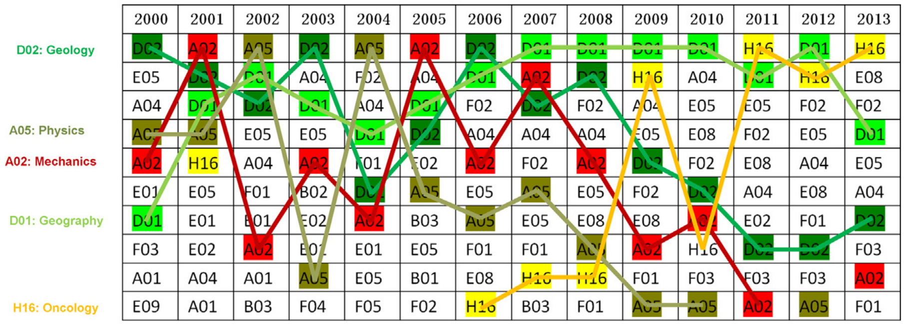

Fig. 8. The top 10 research areas in term of *fundmoney* in each year (the curves represents the changes of ranks of the research areas that ever ranked No.1 from 2000 to 2013 in term of *fundmoney*). Note: The corresponding names of the other codes of research areas are referred to Appendix A.

sponsored money. However, in some basic research areas such as A01: Mathematics obtains a very large number of grants and ranked top 1 in some year, nevertheless, the amount of sponsored money did not remain the same level as the number of grants. Another basic research area A02: Mechanics has better performance in term of the amount of sponsored money than A01: A01: Mathematics. Furthermore, with the development of information technology, the F02: Computer Science obtains more and more number of grants and amount of sponsored money.

4.3. Distributions of funding of research fields

There are 971 number of second-level discipline codes indicating the research field in the NSFC. Distribution fitting is used to select the best statistical distribution for the distributions of funding of research fields. After the goodness of fit tests, it is found that Generalized Pareto distribution is the best fitted function for the dataset. Its Kolmogorov–Smirnov test is around 0.038 and ranked the top one for all the fits across the years.

Probability density function of generalized Pareto distribution is as below. It is specified by three parameters: location μ, scale σ, and shape φ:

$$Pr_t = \frac{1}{\sigma}(1+\varphi z)^{(-(1/\varphi)+1)}, \text{where } z = \frac{x-\mu}{\sigma} \quad (2)$$

In the Generalized Pareto distribution, two parameters, namely shape and scale, is very important to visualize different shapes. Shape is a parameter that must affect the shape of a distribution rather than simply shifting it (as a location parameter does) or stretching/shrinking it (as a scale parameter does). Scale is called a scale parameter, since its value determines the “scale” or statistical dispersion of the probability distribution. If scale is large, then the distribution will be more spread out; if scale is small then it will be more concentrated. The other parameter is location that indicates the shifting on the *x*-axis.

The amount of sponsored money (*fundmoney*) of all the 971 research fields across different years is fitted using Generalized Pareto distribution. The result of fitted parameters is shown in Table 3, and the visualization of distribution fitting is shown in Fig. 9 as well.

The changing of fitted parameters over years is shown in Fig. 10. Since the Scale parameter is increasing as a whole, the distribution become more and more dispersed. It is indicated that some hot research fields obtain more sponsored money, and meanwhile some research areas loss more sponsored money. The top 10 research fields each year in term of *fundmoney* are hot research fields as shown in Appendices B and C. It is shown that the rank of H1617 (Digestive System Neoplasms) increase gradually, and H1602 (Tumorigenesis) and H1606 (Tumor recurrence and metastasis) with top ranks

Table 3
The fitted Generalized Pareto distribution of *fundmoney* of research fields in each year.

k, shape	Delta, scale	mu, location	Year
0.1865	590.29	−23.44	2013
0.21133	516.39	−18.154	2012
0.2767	198.57	−16.279	2011
0.25362	205.88	−24.526	2010
0.22577	125.58	−15.051	2009
0.40712	115.81	−25.539	2008
0.41097	97.052	−21.489	2007
0.391	83.238	−18.131	2006
0.43603	61.024	−13.66	2005
0.53111	47.236	−11.936	2004
0.5428	40.815	−10.424	2003
0.53443	25.788	−6.7571	2002
0.55558	16.771	−4.5332	2001
0.51196	14.334	−3.9738	2000

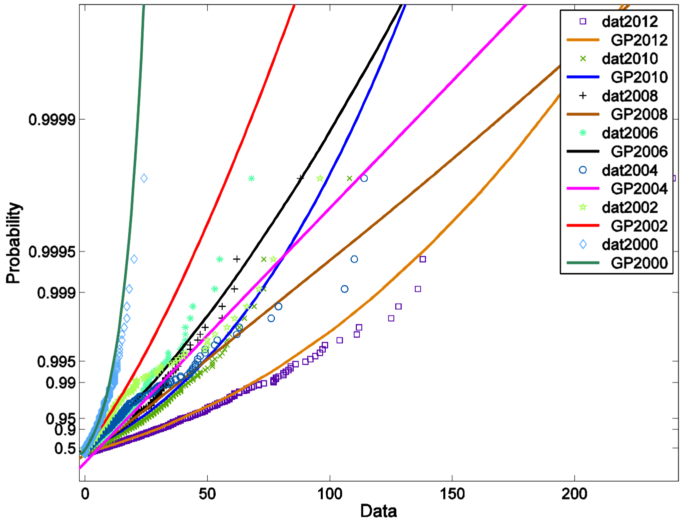

Fig. 9. The P–P plot of the fitted Generalized Pareto distribution of *fundmoney* of research fields in the year of 2000, 2002, 2004, 2008, 2010, 2012.

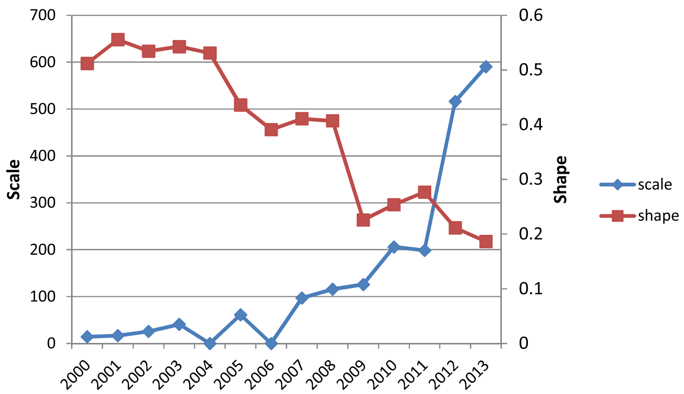

Fig. 10. The shape and scale parameters of fitted Generalized Pareto distribution from 2000 to 2013.

are also related to the tumor of Cancer. In recent years, the research fields in H: Medical sciences and D: Earth sciences are paid more attention to attract a large amount of NSFC funding.

Also, the Shape parameter is decreasing as a whole. This indicates that the tail of the distribution becomes thinner, which means that in the tail of the distribution the convergence speed becomes quicker. In some sense, from 2000 to 2013, the distribution of the amount of sponsored money of research fields becomes more like a Pareto shape, with the more amount of sponsored money obtained by several key supported research areas such as H1617 (Digestive System Neoplasms) and

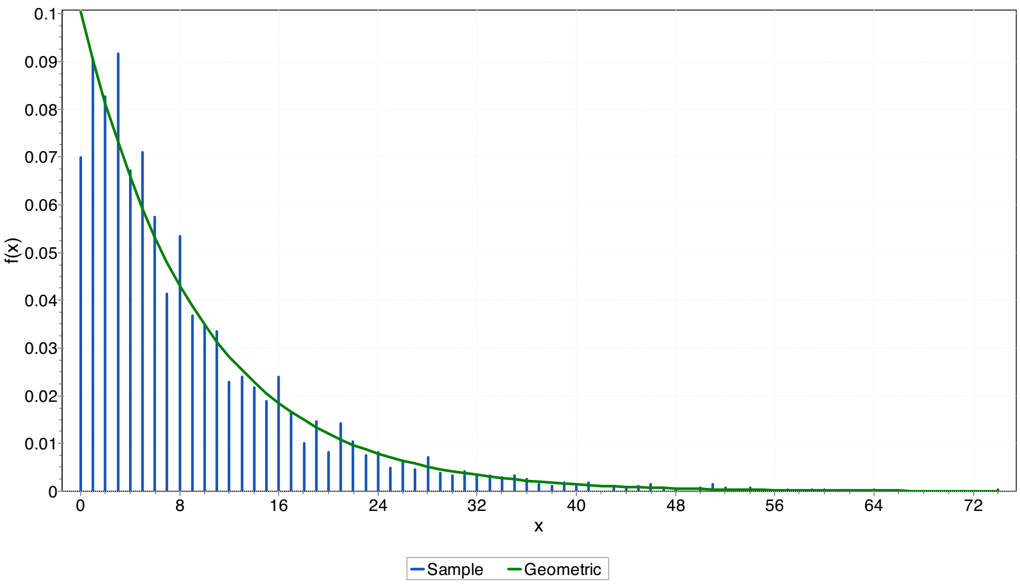

Fig. 11. The fitted Geometric distribution of supported universities of research fields in 2013.

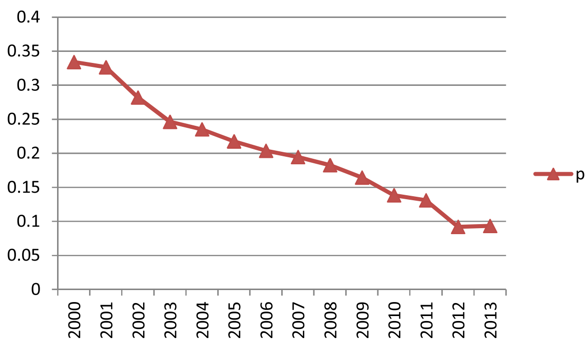

Fig. 12. The probability of success on each trial p of the fitted Geometric distributions over years.

D0309 (Environmental geochemistry). Pareto distribution intrinsically is a power law probability distribution where a few points have very great values and most of points have very small values, which indicates the inequality of *fundmoney* of research fields in nature.

4.4. Distributions of supported universities of research fields

Finally, we also investigate the number of supported universities in each research field of the 971 research fields of the NSFC, because some research field covers many universities but some covers few universities as well. After using the goodness of fit test for many discrete and continual distributions, it is found the Kolmogorov–Smirnov test of Geometric distribution is around 0.075 and ranked the top one for all the years. Therefore, Geometric distribution is the best fitted distribution used to identify and describe the distributions of supported universities of research fields. The visualization of the fitted Geometric distribution in 2013 is shown in Fig. 11 as an example.

Probability density function of Geometric distribution is as below. It is a discrete probability distribution with the parameter of the probability of success on each trial is p, and then the probability that the kth trial (out of k trials) is the first success is:

$$Pr_t = (X = k) = (1 - p)^{k-1} p \quad (3)$$

In the above formula, k denotes the number of supported universities of a certain research field and p is the probability of successfully increasing one unit to attain k in the distribution. We have applied Geometric distribution on the dataset year by year, and the only parameter of the probability p is calculated. The trend of p changing is shown in Fig. 12. It is found the probability of success on the increase of one unit is decreasing as a whole from 2000 to 2013. According to Eq. (3), the probability of attaining the bigger k become larger over years, meaning it is more possible to find a larger number of supported universities for each research field. It also means that in each research field the sponsored money of NSFC covers more universities with a larger possibility. Therefore, the coverage of supported universities in each field has been extended as a whole since 2000.

5. Conclusions

With the development of China's economy, the investment of science and technology has been increased a lot in recent years. Each government sections at the national, province, city level have more and more budget in the scientific research. In this paper, based on one of the most important foundations, National Natural Science Foundation of China (NSFC), we tried to understand the funding allocation policy and the fundamental rule of distributions of funding. The amount of the sponsored money by NSFC has been regarded as a critical indicator for competition among universities each summer when it is the time to publish the results of reviewing NSFC proposals. NSFC announced the summary of sponsored projects in each year, but there is not a whole viewpoint through the investigation of the longitudinal dataset of NSFC. The trend is the key point to summarize the history and make a good policy of scientific funding for the future.

In this paper, we have presented a longitudinal study on the distributions of scientific funding. *First*, among universities, the inequality of funding distribution is checked based on the Gini coefficient (Halffman & Leydesdorff, 2010). Excessive inequality of funding can be unhealthy for academia, as inequality in societies and economies. The finding in China can be as a complement for the recent discussions in the science of inequality (Deaton, 2014). *Second*, distributions of scientific funding across research areas and research fields are investigated as well. *Third*, using distributions fitting techniques, the Generalized Pareto distribution and Geometric distribution is found to describe the distribution of funding and the distribution of supported universities, repetitively, for all the research fields. The investigation of scientific funding will complement the findings from studying bibliometrics dataset (Xie et al., 2014) and help evaluate university research (Geuna & Martin, 2003).

However, this study is not without limitations. Currently, this study has not related the output of the sponsored projects to the input of invested money. From the perspective of policy making, the output performance such as citations and number of publications are more important to evaluate the sponsored projects. With the investigation of outputs of sponsored projects in the future research, the perspective of investigating the allocation of scientific funding will be more complete. In addition, the comparison with other countries and areas such as EU, USA and Japan will benefit more to understand the policy of research funding all over the world.

Acknowledgements

This work was supported by grants from National Natural Science Foundation of China (grant no.: 71373194) and the China Postdoctoral Science Foundation (grant no.: 2014T70143). It was also partly supported by “the Fundamental Research Funds for the Central Universities”.

Appendix A.

86 research areas and their corresponding codes and area numbers.

1:A01	Mathematics	21:C09	Neuroscience, Cognitive Science and Psychology	41:E03	Organic Polymer Materials	61:H06	Motor System
2:A02	Mechanics	22:C10	Biomechanics and Tissue Engineering	42:E04	Metallurgy and Mining	62:H07	Endocrine System/ Metabolic and Nutritional Support
3:A03	Astronomy	23:C11	Physiology and Integrative Biology	43:E05	Mechanical Engineering	63:H08	Hematologic System
4:A04	Physics I	24:C12	Reproductive Biology and Developmental Biology	44:E06	Engineering Thermophysics and Energy Utilization	64:H09	Nervous system And Mental illness
5:A05	Physics II	25:C13	Agronomy Foundation and Crop Science	45:E07	Electrical Science and Engineering	65:H10	Medical Immunology
6:B01	Inorganic Chemistry	26:C14	Plant Protection	46:E08	Built Environment and Structural Engineering	66:H11	Skin and its appendages
7:B02	Organic Chemistry	27:C15	Horticulture and Plant Nutrition	47:E09	Water Conservancy Science and Ocean Engineering	67:H12	Ophthalmology
8:B03	Physical Chemistry	28:C16	Forestry	48:F01	Electronics and Information Systems	68:H13	Otolaryngology Head and Neck Sciences
9:B04	Polymer Science	29:C17	Animal Husbandry and Grassland Science	49:F02	Computer Science	69:H14	Oral Craniofacial Sciences
10:B05	Analytical Chemistry	30:C18	Veterinary Science	50:F03	Automation	70:H15	Severe Medical / Trauma / Burns / Plastic Surgery
11:B06	Chemical Engineering and Industrial Chemistry	31:C19	Fisheries Science	51:F04	Semiconductor Science and Information Device	71:H16	Oncology
12:B07	Environmental Chemistry	32:C20	Food Science	52:F05	Optics and Optoelectronics	72:H17	Rehabilitation Medicine
13:C01	Microbiology	33:D01	Geography	53:G01	Management Science and Engineering	73:H18	Medical Imaging and Biomedical Engineering
14:C02	Botany	34:D02	Geology	54:G02	Business Administration	74:H19	Medical pathogenic microorganisms and infection
15:C03	Ecology	35:D03	Geochemistry	55:G03	Macroeconomic Management and Policy	75:H20	Laboratory Medicine
16:C04	Zoology	36:D04	Geophysics and Space Physics	56:H01	Respiratory System	76:H21	Special Medicine
17:C05	Biophysics, Biochemistry and Molecular Biology	37:D05	Atmospheric Science	57:H02	Circulatory System	77:H22	Radiation Medicine
18:C06	Genetics and Bioinformatics	38:D06	Marine Science	58:H03	Digestive System	78:H23	Forensic Medicine
19:C07	Cytobiology	39:E01	Metallic materials	59:H04	Reproductive System/ Perinatology/ Newborn	79:H24	Endemiology and Occupational Medicine
20:C08	Immunology	40:E02	Inorganic non-metallic materials	60:H05	Urinary System	80:H25	Geriatrics
81:H26	Preventive Medicine	83:H28	Science of Chinese Pharmacology	85:H30	Materia Medica		
82:H27	Traditional Chinese Medicine	84:H29	Chinese and Western Integrative Medicine	86:H31	Pharmacology		

A	Mathematical and physical science	B	Chemical sciences	C	Life sciences	D	Earth sciences
E	Engineering and materials science	F	Information science	G	Management science	H	Medical sciences

Appendix B.

The top 10 research fields in term of *fundmoney* in the year from 2007 to 2013. (The columns in each year represent Discipline code, Name and Sponsored money in ten thousands of Yuan RMB, respectively.)

2013			2012			2011			2010		
H1617	Digestive system neoplasms	11,905	H1617	Digestive system neoplasms	9779.1	D0603	Marine Geology	4665	H1617	Digestive system neoplasms	3206
D0214	Engineering Geology	8879	H1602	Tumorigenesis	7597	H1617	Digestive system neoplasms	4281	E0309	Organic polymer functional materials	3156
H1602	Tumorigenesis	8176.5	D0214	Engineering Geology	6635	H1606	Tumor recurrence and metastasis	4275	C020408	Plant growth and development	2786.8
H1606	Tumor recurrence and metastasis	7455	F010201	Information system modeling and simulation	6225	H1602	Tumorigenesis	3362	C0902	Neurobiology	2394.1
H2902	Integrative Medicine Clinical Foundation	7200	D010105	Cryosphere geography	6124	D0214	Engineering Geology	3062.9	B010303	Functional complex chemical	2333
D0309	Environmental Geochemistry	7006	D0309	Environmental Geochemistry	6101.5	D0309	Environmental Geochemistry	2970.5	H1904	Viruses, virus infection and host immune	2304
D0609	Biological oceanography and Marine biological resources	6658.6	H1606	Tumor recurrence and metastasis	6035.5	C0709	Cell signal transduction	2900	D0507	Climatology and climate system	2161
H2708	Chinese Internal Medicine	6295	H2902	Integrative Medicine Clinical Foundation	5967	D010504	Soil Biology	2571.14	E020301	Advanced structural ceramics	2150
E080402	Sewage treatment and reuse	6108	E0207	Inorganic non-metallic optoelectronic information and functional materials	5905	D0609	Biological oceanography and Marine biological resources	2491	D0609	Biological oceanography and Marine biological resources	2135.3
H0203	Myocardial cell/vascular cell damage, repair, reconstruction and regeneration	5850	H0203	Myocardial cell/vascular cell damage, repair, reconstruction and regeneration	5798.8	H2902	Integrative Medicine Clinical Foundation	2409	D0309	Environmental Geochemistry	2107

2009			2008			2007		
H1617	Digestive system neoplasms	2595	C02	Botany	2690	C07	Cell Biology	2612
F0404	Semiconductor electronic devices	1780	C07	Cell Biology	2652	D0201	Paleontology and paleoecology	2228.4
H1611	Tumour biotherapy	1626	H1617	Digestive system neoplasms	2570	C0601	Plant Genetics	2219
H1904	Viruses, virus infection and host immune	1584	D0507	Climatology and climate system	2388	A0402	Condensed properties II: electronic structure, electrical, magnetic and optical properties	2163.8
D0609	Biological oceanography and Marine biological resources	1464	C05	Biophysics, Biochemistry and Molecular Biology	2219	D0309	Environmental Geochemistry	1823.6
C0606	Gene expression regulation and epigenetics	1434	D0609	Biological oceanography and Marine biological resources	1906	C0702	Cell growth and division	1733
D0214	Engineering Geology	1334	D0601	Physical Oceanography	1894	H1617	Digestive system neoplasms	1665
H2902	Integrative Medicine Clinical Foundation	1303	D0211	Structural geology and active tectonics	1874	C050201	Protein and peptide biochemistry	1644.2
E080402	Sewage treatment and reuse	1274.4	F0202	Computer software	1867	E0207	Inorganic non-metallic optoelectronic information and functional materials	1626.9
D0106	Remote sensing mechanism and method	1254	F0208	Computer network	1856.5	A03	Astronomy	1624.5

Appendix C.

The top 10 research fields in term of *fundmoney* in the year from 2000 to 2006 (The columns in each year represent Discipline code, Name and Sponsored money in ten thousands of Yuan RMB, respectively.)

2006			2005			2004			2003		
D0209	Quaternary Geology	2886	B01	Inorganic Chemistry	3902.6	F050103	Optical storage materials, devices and technology	2097	D0410	Space Physics	1772
F0208	Computer network	2015	E0201	Intraocular lens(IOL)	2016	A050409	New concepts, new principles, new methods	1930	D0507	Climatology and climate system	1763
C050201	Protein and peptide biochemistry	1629	B03	Physical Chemistry	1859.61	F010202	Information Systems Security	1909	D0211	Structural geology and active tectonics	1677
C07	Cell Biology	1594.4	C07	Cell Biology	1732.45	E0805	Structural Engineering	1867.8	E03	Organic polymer materials	1632.77
G0307	Science and technology management and policy	1476.7	H2609	Infectious Disease Epidemiology	1711	B06	Chemical Engineering and Industrial Chemistry	1780.01	E010901	The crystallization and Non-equilibrium solidification of metals	1613
H1611	Tumour biotherapy	1459	C05	Biophysics, Biochemistry and Molecular Biology	1638.35	A0504	Nuclear technology and its applications	1751	E02	Inorganic non-metallic materials	1536.5
H1617	Digestive system neoplasms	1447	C050201	Protein and peptide biochemistry	1340	E0111	Corrosion and Protection of metallic materials	1600	B02	Organic Chemistry	1498.01
D0201	Paleontology and paleoecology	1407.98	A020406	Non-Newtonian flow and rheology	1325	B03	Physical Chemistry	1541.38	B020104	Asymmetric catalysis and reaction	1494
E0207	Inorganic non-metallic optoelectronic information and functional materials	1361.5	E07	Electrical Science and Engineering	1290.4	C130404	Soybean crop germplasm resources and genetic breeding	1527.16	D0201	Paleontology and paleoecology	1456
E0309	Organic polymer functional materials	1342.9	D0601	Physical Oceanography	1237.9	D0201	Paleontology and paleoecology	1521	C1102	Systems physiology	1401

2002			2001			2000		
B05	Analytical Chemistry	1549.9	B05	Analytical Chemistry	1109.3	C07	Cell Biology	464
A05	Physics	1304.81	A05	Physics	1104.75	E0901	Hydrology, water resources	455
E0204	Functional ceramics	925	A020102	Physical mechanics	800	E0904	River coast dynamics and sediment research	430
F010205	Network management	886	C07	Cell Biology	800	B02	Organic Chemistry	420
A050501	Beam physics and accelerator technology	846	B02	Organic Chemistry	783.42	D0201	Paleontology and paleoecology	412
B01	Inorganic Chemistry	796.2	D010106	Integrated Physical Geography	748	B05	Analytical Chemistry	392.6
B03	Physical Chemistry	752.53	H1617	Digestive system neoplasms	705	B03	Physical Chemistry	360
E0804	Environmental Engineering	740.8	C0604	Human Genetics	628	E0805	Structural Engineering	353
D010105	Cryosphere geography	681	B03	Physical Chemistry	618.7	F0301	Control Theory and Methods	348
C050201	Protein and peptide biochemistry	670	D0209	Quaternary Geology	618	E02	Inorganic non-metallic materials	343

References

Abramo, G., Cicero, T., & D'Angelo, C. A. (2013). The impact of unproductive and top researchers on overall university research performance. *Journal of Informetrics*, *7*(1), 166–175.

Auranen, O., & Nieminen, M. (2010). University research funding and publication performance—An international comparison. *Research Policy*, *39*(6), 822–834.

Chatterjee, S., & Hadi, A. S. (2012). *Regression analysis by example*. John Wiley & Sons.

Choi, S. (2010). Globalization, China's drive for world-class universities (211 Project) and the challenges of ethnic minority higher education: The case of Yanbian university. *Asia Pacific Education Review*, *11*(2), 169–178.

Deaton, A. (2014). Inevitable inequality? *Science*, *344*(6186), 783. http://dx.doi.org/10.1126/science.1255661

Dorfman, R. (1979). Formula for the Gini coefficient. *Review of Economics and Statistics*, 61.

Geuna, A., & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva*, *41*(4), 277–304.

Halffman, W., & Leydesdorff, L. (2010). Is inequality among universities increasing? Gini coefficients and the elusive rise of elite universities. *Minerva*, *48*(1), 55–72.

Lane, J. (2009). Assessing the impact of science funding. *Science*, *324*(5932), 1273.

Lok, C. (2010). Science funding: Science for the masses. *Nature*, *465*(7297), 416.

Payne, A. A., & Siow, A. (2003). Does federal research funding increase university research output? *Advances in Economic Analysis & Policy*, *3*(1).

Pendlebury, D. A. (1991). Science, citation, and funding. *Science*, *251*(5000), 1410–1411.

Roebber, P. J., & Schultz, D. M. (2011). Peer review, program officers and science funding. *PLoS ONE*, *6*(4), e18680.

Shibayama, S. (2011). Distribution of academic research funds: A case of Japanese national research grant. *Scientometrics*, *88*(1), 43–60.

Wu, Y. H. (2013). The cross-state distribution of federal funding in the USA: The case of financing academic research and development. *Science and Public Policy*, *40*(3), 316–326.

Xie, Y. (2014). Undemocracy: Inequalities in science. *Science*, *344*(6186), 809–810.

Xie, Y., Zhang, C., & Lai, Q. (2014). China's rise as a major contributor to science and technology. *Proceedings of the National Academy of Sciences of the United States of America*, http://dx.doi.org/10.1073/pnas.1407709111

Zhang, H., Patton, D., & Kenney, M. (2013). Building global-class universities: Assessing the impact of the 985 Project. *Research Policy*, *42*(3), 765–775.