



Using context-dependent DEA to analyze the efficiency of highly funded scientists in China

Keyu Xiang¹ · Haiming Liang¹ · Zhaoxia Guo¹ · Yucheng Dong¹

Received: 9 April 2021 / Accepted: 19 July 2021 / Published online: 31 July 2021
© The Author(s) 2021

Abstract

Funding inputs and research outputs have always been two central issues in the science of science. In recent decades, research funding plays an increasingly important role in scientific research. Thus, it is progressively significant for management authorities to measure the research efficiency of highly funded scientists, which can be helpful for them to make effective policies. However, few researchers use quantitative analysis to study these issues. To promote the research in this field, we begin with collecting a dataset. This dataset contains research funding and other information from 345 highly funded scientists in Mainland China. Next, we use the dataset to measure the efficiency of highly funded scientists based on the data envelopment analysis. In this way, highly funded scientists are placed into several levels according to their research inputs and outputs. We also give their attractiveness and progress scores compared to other grades. The learning path for less efficient scientists is also provided. We find that highly funded scientists have relatively high efficiency in three kinds of projects, such as the Major Research Plan. Besides, the career length and career start year are demonstrated to have a limited impact on the highly funded scientists. These patterns are beneficial for the development of the scientific community and management authorities to make policies.

Keywords Data envelopment analysis · Highly funded scientists · R&D efficiency · Research funding · Research outputs

Introduction

In science of science, research efficiency has always been a hot topic. Applications of the formal theories of system science in science of science are demonstrated, which reveals how system structural and behavioral complexities may be efficiently reduced [28]. As a popular and effective mathematical method, data envelopment analysis (DEA) has been deeply studied and widely used in many fields [2, 22]. Bagheri et al. [2] proposed a DEA-based method to multi-objective shortest path problem. Pakkar and Mohammad Sadegh [22] integrated DEA and analytic hierarchy process (AHP) approach to obtain attribute weights.

Recently, researchers favor using the DEA to estimate a functional relationship between multiple research inputs and outputs rather than sum weighted quantitative indicators directly. Lee et al. [15] evaluated Korea's R&D investment

performance at the local government level using slack-based model data envelopment analysis (SBM-DEA). Qin et al. [24] measured regional R&D efficiency and its spillover effects in China using DEA as a performance analysis tool. Zemtsov et al. [30] applied DEA to estimate the efficiency in creating technologies of different regions in Russia. Guoliang et al. [29] used DEA to analyze the R&D efficiency issues between countries. They analyzed R&D efficiency issues from the perspectives of countries and regions. These studies prove the effectiveness of the DEA method on the topic of R&D efficiency. At the same time, intuitively speaking, it is not feasible to add up quantitative indicators through weights to study the problem of R&D efficiency. We cannot simply add up the data at two different dimensions by weight, where the interpretability is also very poor.

In recent decades, with the shortage of scientific funding and the increasing importance of scientific funds for scientific research, those highly funded researchers always catch more interest and notice from others than normal scientists [11]. Some researchers have found that even famous universities may be inefficient in publishing papers [31]. Thus, we are curious whether highly funded scientists tend

✉ Zhaoxia Guo
zx.guo@alumni.polyu.edu.hk

¹ Business School, Sichuan University, Chengdu 610064, China

towards making more worthwhile research achievements. Yet, quantitative research regarding the R&D efficiency of highly funded scientists is still a blank. The concerning dataset of funding inputs and research paper outputs, providing the possibility of quantitative research, has not been publicly reported so far, either.

Thus, this paper seeks to measure the efficiency in using the research funding resources of highly funded scientists based on DEA. As one of the most significant indicators for measuring scientific outputs, research papers have been widely used to evaluate scientists' achievements. Because there is no relevant public dataset that can satisfy our analysis, we first collect a dataset concerning research funding and other information of highly funded scientists in Mainland China. The dataset obtained is specifically targeted at the basic personal information of highly funded scientists, and their research projects funded by the National Natural Science Foundation of China (NSFC), and research papers supported by these projects. Three sorts of sets of data, including scientists' R&D funding, the number of citations, and publications supported by these NSFC projects, have been extracted from the dataset for analysis as simply described in the following.

Since scientists in different academic divisions perform differently in the number of papers and citations [14, 21], we separately use the DEA to analyze scientists in their specific academic divisions. In this way, scientists in the same academic division are graded into several tiers according to their R&D efficiency. Every less efficient scientist is provided a gradual learning path for improving their efficiency. In addition, their attractiveness and progress scores compared to other tiers are well shown. On this basis, the R&D efficiency of scientists in their specific academic division is well estimated.

To further investigate highly funded scientists' R&D efficiency, we analyze the R&D efficiency of highly funded scientists on different types of NSFC projects. We find that the ten different kinds of NSFC projects are graded into four levels. Three sorts of projects are classified into the first level. The General Program project has the biggest progress scores compared to the other context levels, which means that highly funded scientists perform badly in this type of project.

To analyze whether the career stage affects highly funded scientists' research efficiency, we conduct the following analysis. First, we use the k-means method [20] to clustering scientists into nine clusters based on their career length and career start year. Then their R&D efficiency is analyzed through the DEA models. We find that career length and career start year have a limited effect on highly funded scientists' efficiency, with scientists who have relatively short careers being slightly more efficient.

These results measure the R&D efficiency of highly funded scientists from multiple perspectives benefiting the effective policy-making for relevant management authorities.

Methods

Input and output indicators

Our dataset, which has not been made public, contains the effective NSFC funding and the information of papers supported by those NSFC projects of 345 highly funded scientists in Mainland China until July 2020, which can satisfy our analysis needs. It is worth noting that these 345 highly funded scientists are selected based on the total amount of funding applied by each academic division from 1986 to 2020. Due to the long duration of some large projects, these projects may not be completed until we collect the data. Therefore, we only consider the completed projects. Since there are few projects left, after removing these large projects, for some young and highly funded scientists, the analytical sense of these scientists may be reduced. Therefore, we removed some of the scientists manually and did other related processing to get this dataset. The dataset will be published in data journals soon.

Therefore, we select the research funding as our research input indicator, which is one of the most popular indexes for evaluating scientists' scientific resources. The number of papers and citations of the papers supported by the research funding mentioned above are selected as the output indicators. Both of them, which can be easily extracted from our dataset, are common indicators that are used for evaluating the R&D efficiency scientists. Besides, before using DEA to analyze, we manually remove three scientists (i.e., a30, a163, a310) from the dataset, whose main researches may produce few papers, such as the study of the Moon, which would deeply affect the results if we do not remove these data. For better estimating the R&D efficiency, several sets of data are extracted from our dataset. At last, we totally obtain eleven sets of data for analysis. Eight of them are extracted based on the academic division, which are used to estimate the efficiency of individual scientists in their academic division; one set of data is extracted based on the NSFC project type to measure highly funded scientists' preference on different types of NSFC projects; one set is extracted based on scientists' career length and their career start year to measure the difference of R&D efficiency in scientists who have distinct career length. To succinctly show our input and output indicators, we present the set of data extracted based on the NSFC project type. See Table 1 for details. We should note that the unit of research funding is ten thousand in CNY.

Table 1 Values of input and output indicators across ten different types of NSFC funding projects

DMU	Research funding	Papers	Citations
National science fund for distinguished young scholars	23,169.90	3573	140,807
Major program	5693.00	532	13,401
Joint funds	82,937.70	12,228	341,288
Young scientists fund	186,502.90	8831	217,589
Key program projects	35,866.00	3694	116,650
Fund for less developed regions	31,105.00	2449	70,928
Major research plan	168,747.50	27,299	823,203
General program	146,570.00	3626	48,072
Special fund for research on National Major Research Instruments	19,575.02	837	22,193
Science fund for creative research groups	886.90	65	6468

Classical DEA models

Since Charnes, Cooper and Rhodes (CCR) first introduced the DEA method in 1978 [10], it had been used to analyze the efficiency in many fields [1, 8, 12, 26, 29]. This model could measure the efficiency of decision-making units (DMUs) under the assumption of constant returns to scale (CRS). Then, some scientists argued the returns to scale (RTS) were variable sometimes. Banker, Charnes and Cooper (BCC) proposed the BCC model in 1987 [3]. Here, we present these two models briefly as follows:

Let $X_k = (x_1, \dots, x_m)$, $Y_k = (y_1, \dots, y_s)$ be input and output vectors of m and s dimensions, separately, of $DMU_k (k = 1, 2, \dots, n)$, and $\phi^*(k)$ be the optimal value which represents the efficiency of DMU_k . We have the following input-based CCR model [10]:

$$\begin{aligned} \phi^*(k) = \min \phi(k) \\ \text{s.t.} \begin{cases} \sum_{i=1}^n \lambda_i X_i \leq \phi(k) X_k \\ \sum_{i=1}^n \lambda_i Y_i \geq Y_k \\ \lambda_i \geq 0, i = 1, \dots, n, \end{cases} \end{aligned} \tag{1}$$

where $\lambda_i \geq 0$ are the multipliers of inputs and outputs. The size of $\phi(k)$ ranges from 0 to 1.

If we assume that the production technology satisfies variable returns to scale (VRS) assumption, we have the following input-based BCC model [3]:

$$\begin{aligned} \phi^*(k) = \min \phi(k) \\ \text{s.t.} \begin{cases} \sum_{i=1}^n \lambda_i X_i \leq \phi(k) X_k \\ \sum_{i=1}^n \lambda_i Y_i \geq Y_k \\ \sum_{i=1}^n \lambda_i = 1 \\ \lambda_i \geq 0, i = 1, \dots, n. \end{cases} \end{aligned} \tag{2}$$

Context-dependent DEA model

The context-dependent DEA model was proposed by Seiford and Zhu in 2003 to grade DMUs into multi-levels, assuming that the production technology satisfies the CRS assumption [27]. Inspired by this model, Guoliang et al. [2] proposed to grade DMUs, assuming the production technology satisfies the VRS assumption [29]. The following algorithm explains the procedure:

Step1: Run the DEA model (2) to obtain a production frontier.

Step2: Exclude the DMUs in the above production frontier, and run the DEA model (2) for the remaining DMUs again to obtain another production frontier.

Step3: Repeat the above two steps until no DMU is left.

Let $J^1 = \{DMU_j, j = 1, \dots, n\}$ be the set of all n DMUs. We define $J^{l+1} = J^l - E^l$, where $E^l = \{DMU_q \in J^l | \text{efficientscore} = 1\}$. In this way, we are given several production frontiers grading DMUs into different grades. Assuming that we divide the DMUs into L production frontiers, we can measure the attractiveness of the DMUs through the following model:

$$\Omega_q^*(d) = \min \Omega_q(d), \quad d = 1, \dots, L - 1$$

$$\text{s.t.} \begin{cases} \sum_i \lambda_i X_i \leq \Omega_q(d) X_q \\ \sum_i \lambda_i Y_i \geq Y_q \\ \lambda_i \geq 0, i \in J^D, D = d + 1, \dots, L, \end{cases} \quad (3)$$

where the J^D is the DMU set in frontier D . $\Omega_q^*(d)$ is the attractiveness index for the DMU_q in frontier d , which is evaluated by context frontier D . In other words, we are given a series of attractiveness scores of DMU_q evaluated by less efficient context frontier. In this way, we compare every DMU in frontier D with the lower performance context frontiers and distinguish the efficiency of DMUs in the same frontier through their attractiveness score to other lower performance efficient DMUs.

Likewise, under the assumption of the VRS, we can measure the progress of the DMUs through the following model:

Table 2 The descriptive statistics resulted from the dataset

Statistics	Academic division							
	Mathematical and physical sciences	Chemical sciences	Life sciences	Earth sciences	Engineering and materials science	Information sciences	Management sciences	Health sciences
Number of scientists	42	46	49	46	46	44	37	37
Number of projects	280	457	418	362	393	267	243	251
Number of projects per scientist	6.67	9.93	8.53	7.87	8.54	6.07	6.57	6.78
Average funding amount per scientist (10 k CNY)	2536.46	2808.90	2227.18	2190.76	1787.44	2174.50	635.94	1765.77
Average number of publications per project	17.30	31.40	11.38	12.38	22.82	23.56	10.66	14.18
Average number of publications per million CNY	4.55	11.11	4.36	4.45	10.91	6.57	11.01	5.45

$$\begin{aligned}
 P_q^*(g) &= \min P_q(g), \quad g = 2, \dots, L \\
 \text{s.t.} &\left\{ \begin{aligned}
 \sum_i \lambda_i X_i &\leq P_q(g) X_q \\
 \sum_i \lambda_i Y_i &\geq Y_q \\
 \sum_i \lambda_i &= 1 \\
 \lambda_i &\geq 0, i \in J^G, G = 1, \dots, g - 1,
 \end{aligned} \right. \quad (4)
 \end{aligned}$$

where the J^G is the set of DMUs in the frontier G . $B_q^*(g) = 1/P_q^*(g)$ is called the progress of DMU $_q$ in the frontier g , which is evaluated by context frontier G . The reference set of DMU $_k$ in the frontier g based on the context G for $G < g$ is given by

$$R_k^{CD}(G) = \{i \in J^G | \lambda_i > 0 \text{ in Eq. (4)}\}. \quad (5)$$

Results

Descriptive statistics

Table 2 presents the descriptive statistics resulting from the dataset in terms of different academic divisions. We find that scientists' funding and productivity vary across different academic divisions. The scientists in the academic division of chemical sciences are funded the largest number of research projects, whereas the scientists in the information sciences division are funded the least on average. We also find that the average funding amount per scientist in the academic division of management sciences is much less than the average amount in other academic divisions.

Conduct separate efficiency analysis of highly funded scientists from different academic divisions

The structure of R&D produce modes in different academic divisions, such as the amount of research funding and the number of papers supported by the funding, would behave very differently [9, 25]. Therefore, we decide to divide highly funded scientists based on their academic divisions. Then, we analyze these scientists in the same academic division. We believe that, in this way, the research efficiency of scientists will be appropriately evaluated.

Performance level

Take the scientists in the chemical sciences academic division as an example for detailed analysis. All the 46 highly funded scientists in the chemical sciences academic division

are graded into nine grades according to their R&D efficiency. They are

- $E^1 = \{a48, a49, a53, a57, a66\}$,
- $E^2 = \{a45, a46, a47, a65, a68, a72, a78, a84, a85, a86\}$,
- $E^3 = \{a55, a56, a67, a76, a83, a87, a88\}$,
- $E^4 = \{a44, a51, a52, a58, a59, a62, a70, a79\}$,
- $E^5 = \{a54, a60, a64, a71, a77, a80, a81\}$,
- $E^6 = \{a50, a75, a82\}$,
- $E^7 = \{a61, a73, a74\}$,
- $E^8 = \{a63, a69\}$,
- $E^9 = \{a43\}$.

The efficient score of a48, a49, a53, a57, and a66, is one. Thus, they establish the first efficiency frontier. Apart from these five scientists, the efficiency score of ten scientists in the remaining 41 scientists is one. Thus, they establish the second efficiency frontier. Similarly, to find another efficiency frontier, we exclude these ten scientists. At last, the other scientists are assigned distinct efficiency frontiers. Obviously, a43 is the only scientist on the ninth efficiency frontier, who can, thus, be considered the worst performer in the chemical sciences academic division. It is not hard to observe that we can use the production frontier of the BCC-DEA model multiple times to obtain multiple efficiency frontiers, which can place all scientists at different performance levels. Moreover, every scientist, who is placed at the less efficient level and wants to improve his R&D efficiency, can easily find an appropriate and gradual learning path containing a set of scientists in the high-efficiency frontiers.

Reference set

Table 3 presents the reference sets of each scientist in the chemical sciences academic division based on model (5). Scientist a43 is classified into Level 4. His reference sets consist of a56, a76, and a88 in Level 3, a45, a65, and a68 in Level 2, a53, and a66 in Level 1. Then he expects to boost his performance on the R&D efficiency, which may give him more chance to apply for research funding. Directly picking a53 and a66 at Level 1 as the learning objects is probably not the best choice. This is because the length between these two levels is so considerable.

In contrast, he can consider the reference set at Level 3. It is a suitable learning target for him. As for which scientists we should choose from each reference set to learn, we can find the answer based on every scientist's attractiveness and progress rank in their own level. See details in the following section.

Table 3 Reference sets for scientists in chemical sciences academic division by context-dependent DEA

Scientists code	Level	$R_k^{CD}(1)$	$R_k^{CD}(2)$	$R_k^{CD}(3)$	$R_k^{CD}(4)$	$R_k^{CD}(5)$	$R_k^{CD}(6)$	$R_k^{CD}(7)$	$R_k^{CD}(8)$
a43	9	a57, a66	a84	a67, a87	a70	a71	a82	a73	a63
a44	4	a53, a66	a45, a65, a68	a56, a76, a88					
a45	2	a53, a57							
a46	2	a53, a66							
a47	2	a53, a57							
a48	1								
a49	1								
a50	6	a53, a66	a45, a65, a68	a76, a87	a58, a70	a80, a81			
a51	4	a53, a57	a45, a68	a56, a83					
a52	4	a53, a66	a45, a65, a68	a55, a76, a83					
a53	1								
a54	5	a53, a57	a45, a68	a56, a83	a51, a58, a79				
a55	3	a53, a66	a45, a86						
a56	3	a53, a57, a66	a45, a65, a68						
a57	1								
a58	4	a53, a57, a66	a45, a65, a68	a76, a87, a88					
a59	4	a53, a57	a45, a68	a56, a83					
a60	5	a53, a57	a45, a68	a87, a88	a51, a58, a70				
a61	7	a53, a57	a45, a65, a68	a76, a87	a58, a70	a71, a80, a81	a75, a82		
a62	4	a53, a66	a45, a65, a85	a76, a87					
a63	8	a53, a57	a45, a68	a67, a87	a58, a70	a71, a81	a75, a82	a61, a73	
a64	5	a53, a57	a45, a68	a56, a88	a51, a58, a79				
a65	2	a53, a66							
a66	1								
a67	3	a57, a66	a84						
a68	2	a53, a57							
a69	8	a53, a57	a45, a68	a76, a87	a58, a70	a71, a80, a81	a75, a82	a61, a73	
a70	4	a53, a57	a68, a84	a67, a87					
a71	5	a53, a57	a45, a68	a67, a87	a51, a70				
a72	2	a53, a66							
a73	7	a53, a57	a68, a84	a67, a87	a70	a71	a82		
a74	7	a53, a57	a45, a68	a87, a88	a51, a70	a60, a71	a75, a82		
a75	6	a53, a57	a45, a68	a87, a88	a51, a58, a70	a60, a71			
a76	3	a53, a66	a45, a65, a85						
a77	5	a53, a57	a45, a65, a68	a56, a88	a58, a79				
a78	2	a53, a57							
a79	4	a53, a57	a45, a68	a56, a83					
a80	5	a53, a66	a45, a65, a68	a76, a87	a58, a79				
a81	5	a53, a66	a45, a65, a68	a76, a87	a58, a70				
a82	6	a53, a57	a68, a84	a67, a87	a58, a70	a71, a81			
a83	3	a53, a57	a45, a68						
a84	2	a53, a57							
a85	2	a53, a66							
a86	2	a53, a57							
a87	3	a53, a57	a45, a68						
a88	3	a53, a57	a45, a68						

Table 4 Attractiveness and Progress scores for 46 scientists in chemical sciences academic division by context-dependent DEA

	Scientists code	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9
Level 1	a48		1.000(5)	1.000(5)	1.637(5)	1.715(5)	3.469(5)	4.607(5)	7.294(5)	79.146(5)
	a49		1.134(2)	1.646(2)	2.064(2)	2.875(2)	4.376(2)	5.862(2)	9.200(2)	101.719(2)
	a53		1.278(1)	1.812(1)	2.491(1)	3.211(1)	5.279(1)	7.011(1)	11.100(1)	120.443(1)
	a57		1.000(5)	1.203(3)	1.654(4)	2.132(3)	3.505(4)	4.654(4)	7.370(4)	79.964(4)
	a66		1.000(5)	1.083(4)	1.802(3)	1.888(4)	3.819(3)	5.071(3)	8.030(3)	87.126(3)
Level 2	a45	1.210(3)		1.457(1)	1.932(3)	2.567(1)	4.096(3)	5.439(3)	8.612(3)	93.444(3)
	a46	5.429(10)		1.000(10)	1.000(10)	1.000(10)	1.000(10)	1.286(9)	2.037(9)	22.102(9)
	a47	5.223(9)		1.000(10)	1.000(10)	1.000(10)	1.000(10)	1.213(10)	1.800(10)	21.142(10)
	a65	1.029(1)		1.402(3)	2.331(1)	2.443(3)	4.940(1)	6.560(1)	10.388(1)	112.708(1)
	a68	1.771(6)		1.000(10)	1.000(10)	1.284(7)	1.809(7)	2.639(7)	3.910(7)	46.106(7)
	a72	1.804(7)		1.000(10)	1.372(6)	1.452(6)	2.908(6)	3.861(6)	6.114(6)	66.336(6)
	a78	1.355(5)		1.309(4)	1.590(5)	2.272(4)	3.336(5)	4.659(5)	7.014(5)	81.130(5)
	a84	1.985(8)		1.000(10)	1.000(10)	1.035(8)	1.178(8)	2.117(8)	3.204(8)	37.776(8)
	a85	1.153(2)		1.273(5)	2.118(2)	2.219(5)	4.489(2)	5.961(2)	9.438(2)	102.404(2)
a86	1.240(4)		1.435(2)	1.846(4)	2.517(2)	3.913(4)	5.196(4)	8.227(4)	89.260(4)	
Level 3	a55	3.101(6)	2.372(7)		1.000(7)	1.000(7)	1.686(4)	2.239(5)	3.546(5)	38.471(6)
	a56	1.691(3)	1.278(4)		1.430(2)	1.784(2)	3.032(2)	4.026(2)	6.375(2)	69.165(2)
	a67	3.323(7)	1.319(5)		1.000(7)	1.000(7)	1.000(7)	1.022(7)	1.558(7)	18.369(7)
	a76	1.472(1)	1.217(3)		1.663(1)	1.798(1)	3.525(1)	4.681(1)	7.411(1)	80.416(1)
	a83	2.768(5)	2.273(6)		1.000(7)	1.084(5)	1.449(6)	2.229(6)	3.316(6)	39.097(5)
	a87	1.870(4)	1.148(1)		1.000(7)	1.295(4)	1.486(5)	2.667(4)	4.009(4)	47.277(4)
	a88	1.622(2)	1.197(2)		1.179(3)	1.722(3)	2.087(3)	3.547(3)	5.314(3)	62.667(3)
	a44	2.892(5)	2.207(5)	1.737(7)		1.026(4)	1.781(4)	2.366(5)	3.746(5)	40.641(5)
Level 4	a51	1.934(1)	1.492(1)	1.072(2)		1.479(1)	1.684(5)	3.040(1)	4.579(1)	53.990(1)
	a52	3.692(8)	2.921(8)	1.501(6)		1.000(8)	1.405(6)	1.866(6)	2.954(6)	32.054(6)
	a58	2.405(3)	1.654(3)	1.351(3)		1.173(3)	1.934(2)	2.568(3)	4.067(3)	44.125(3)
	a59	3.306(6)	2.621(7)	1.480(4)		1.000(8)	1.080(7)	1.818(7)	2.722(7)	32.102(7)
	a62	2.874(4)	2.466(6)	1.866(8)		1.000(8)	1.795(3)	2.384(4)	3.774(4)	40.954(4)
	a70	3.596(7)	1.942(4)	1.501(6)		1.000(8)	1.000(8)	1.225(8)	1.866(8)	22.007(8)
	a79	2.064(2)	1.607(2)	1.057(1)		1.429(2)	2.120(1)	2.925(2)	4.457(2)	50.887(2)
	a54	4.126(7)	3.167(7)	2.392(7)	2.043(7)		1.000(7)	1.435(6)	2.140(6)	25.237(6)
	a60	2.675(2)	1.861(1)	1.577(3)	1.121(3)		1.171(5)	2.061(2)	3.095(3)	36.500(3)
Level 5	a64	2.878(3)	2.170(3)	1.728(5)	1.355(5)		1.454(3)	2.053(3)	3.056(4)	35.789(4)
	a71	3.883(6)	2.288(5)	1.891(4)	1.214(4)		1.000(7)	1.221(7)	1.861(7)	21.940(7)
	a77	3.336(5)	2.533(6)	1.978(6)	1.494(6)		1.416(4)	1.881(5)	2.978(5)	32.312(5)
	a80	2.507(1)	1.883(2)	1.434(1)	1.005(1)		2.022(1)	2.686(1)	4.252(1)	46.139(1)
	a81	2.982(4)	2.204(4)	1.528(2)	1.102(2)		1.672(2)	2.220(4)	3.514(2)	38.133(2)
	a50	6.281(3)	4.788(3)	3.286(3)	2.389(3)	2.175(3)		1.057(2)	1.674(2)	18.165(2)
	a75	3.035(1)	2.101(1)	1.782(1)	1.261(1)	1.121(2)		1.808(1)	2.719(1)	32.057(1)
	a82	4.540(2)	2.509(2)	1.955(2)	1.290(2)	1.029(1)		1.000(3)	1.502(3)	17.715(3)
Level 6	a61	5.187(2)	3.230(1)	2.615(1)	1.756(1)	1.494(2)	1.305(1)		1.583(1)	17.339(2)
	a73	6.908(3)	3.570(3)	2.718(2)	1.808(2)	1.431(1)	1.388(2)		1.000(3)	11.081(3)
	a74	5.088(1)	3.248(2)	2.806(3)	1.825(3)	1.553(3)	1.430(3)		1.524(2)	17.969(1)
Level 7	a63	7.885(2)	4.534(2)	3.632(2)	2.344(2)	1.906(2)	1.797(2)	1.301(2)		10.543(2)
	a69	7.291(1)	4.339(1)	3.491(1)	2.262(1)	1.888(1)	1.706(1)	1.274(1)		11.792(1)
Level 8	a43	50.118(1)	17.106(1)	12.948(1)	8.588(1)	6.794(1)	6.592(1)	4.749(1)	3.539(1)	

Table 4 presents the attractiveness and progress scores of 46 scientists in the chemical sciences academic division. The bold number represents progress scores, and the black number represents attractiveness scores. It is easy to observe that the diagonal line is the comparison of scientists at the same level, the values of which are definitely one. Therefore, we leave the blank here to make the whole table beautiful and easy to observe. It is worth noting that the bracket at the right of the number indicates the ranking position of the scientists at the same level, which is given based on the attractiveness or progress scores. A smaller number means a better rank. The higher attractiveness of a distinct scientist indicates an ample space from a certain low-performance level. In other words, it means that the scientist performs better than the other scientists at the same frontier.

Similarly, the lower progress score indicates a small space from the guiding level. In other words, it means that the scientist needs the slightest effort to keep up with the scientists who show high efficiency. The performance of scientists in the same level can be easily discerned through the rank. As shown in Table 4, the attractiveness of scientists at the guiding frontier compared to scientists at laggard levels and the progress of scientists at the laggard level compared to guiding levels both show a gradual upward trend.

In Level 1, a53 is the most attractive and scientist a48 is the least attractive. Every scientist wants the ideal situation, that is higher attractiveness, but lower progress scores. Both a45 and a65 do a good job, who have a relatively high attractiveness score and a relatively low progress score in Level 2. Nevertheless, a46 has a bad performance on attractiveness and progress scores and is, therefore, one of the worst performers in Level 2. Besides, it is not hard to observe that scientist a47 behaves as severely as a46.

Moreover, some scientists show high attractiveness scores at distinct levels. We also observed that some scientists' rankings of attractiveness and progress score fluctuated, such as the a85 in Level 2. This may be due to the fact that the scientific production mode of a certain level scientist is not similar to that of this scientist. By observing this scientist's overall attractiveness and progress scores at all levels, we can judge whether this scientist is excellent enough and whether the scientific research and mode of this scientist are worth learning. The R&D efficiency of all scientists in the chemical academic division at other Levels can also be well discerned from Table 4.

Implication

As described in Table 3, the reference sets of distinct scientists can consist of over one scientist. The efficiency ranking

of scientists can help less efficient scientists find more suitable learning targets. For example, as shown in Table 3, a55 gets a reference set at Level 2, which consists of a45 and a86. In order to get the most appropriate target to learn at this level, after measuring the scores of these two scientists in Table 4, we consider that a45 performs better, because of his high ranks. By this rule, it is speculated that the appropriate learning path of a55 from a low-performance level to a high-performance level should be a45, a53. Thus, scientists at the lagging level would be suggested an appropriate and gradual learning path to boost their R&D efficiency.

The progress scores of the scientists show that the bigger the progress score of a scientist, the more considerable effort the scientist needs to keep up with the scientists who show high efficiency. Similarly, the smaller the progress score of a scientist, the less effort the scientist needs to keep up with the scientists who show high efficiency.

On that wise, we grade scientists into multiple levels according to their R&D efficiency. Moreover, the R&D efficiency distance of every scientist compared to other levels is well described through the attractiveness and progress scores. We can even discern the difference between scientists in the same level. Besides, a gradual and relatively appropriate learning path for less efficient scientists to improve their performance is provided. These results can help scientists understand their R&D efficient position in the scientific community and the scientific funding competitive environment. Relative management authorities may make more effective policies to improve the R&D efficiency of scientists on the basis of these facts.

Efficiency analysis of different types of funding projects

Performance level

To check the R&D efficiency of highly funded scientists in different types of NSFC projects, we first consider the efficient frontier of the CCR-DEA model to grade these projects. Three types of projects are graded into the first level, and the other seven types of projects are evenly graded into other levels. In other words, each level only contains one type of project. Obviously, this model does not consider the effect of the funding scale and performs not well enough on efficiency measuring of the NSFC funding project. Thus, we choose to use the efficiency frontier of the BCC-DEA model to measure the efficiency of multiple types of NSFC projects. We grade ten different types of funding into four grades. They are

Table 5 Reference sets for NSFC projects by context-dependent DEA

NSFC funding project code	Level	$R_k^{CD}(1)$	$R_k^{CD}(2)$	$R_k^{CD}(3)$
P1	1			
P2	1			
P3	1			
P4	2	P2, P3		
P5	2	P2, P3		
P6	3	P2, P3	P4, P5	
P7	3	P1, P2, P3	P4, P5	
P8	3	P2, P3	P4, P5	
P9	4	P2, P3	P4, P5	P7, P8
P10	4	P2, P3	P4, P5	P7, P8

Table 6 Attractiveness and Progress scores for ten types of NSFC projects by context-dependent DEA

	NSFC funding project code	Level 1	Level 2	Level 3	Level 4
Level 1	P1		1.477(2)	1.869(2)	2.665(2)
	P2		1.185(3)	1.571(3)	2.139(3)
	P3		1.772(1)	2.242(1)	3.198(1)
Level 2	P4	1.512(2)		1.000(2)	1.187(2)
	P5	1.093(1)		1.431(1)	1.873(1)
Level 3	P6	3.396(2)	3.083(3)		1.000(3)
	P7	1.540(1)	1.195(1)		1.426(1)
	P8	3.468(3)	2.521(2)		1.000(3)
Level 4	P9	1.996(1)	1.616(1)	1.081(1)	
	P10	6.418(2)	5.610(2)	4.131(2)	

BCC-DEA model’s production frontier multiple times, we can yield several efficiency frontier lines to place all types of projects into different performance levels. In addition, relative management authorities can find a series of reference set for less efficient projects to improve their efficiency.

Reference set

Table 5 shows the reference sets of each NSFC project based on model (5). P9 is classified at Level 4. Its reference sets consist of P7 and P8 at Level 3, P5 and P4 at Level 2, P2 and P3 at Level 1. Then, if relative management authorities want to improve the efficiency of this type of project, they can consider the reference set at Level 3 as practical and gradual learning aim to reduce the change of resource distribution and see whether the result is satisfying.

But, which NSFC projects should we choose from each reference set to let the project learn? We can refer to the scores of every project shown in Table 6. The higher attractiveness of a distinct project type indicates an ample space from a low-performance frontier, which means that this type of project performs better than the other types of projects at the same frontier. Similarly, the lower progress score indicates a small space from the guiding level. It means that this type of project needs the least resources to keep up with the NSFC projects that show high efficiency. Therefore, we always select the project which shows high attractiveness and low progress ranks.

Attractiveness and progress

Table 6 shows the attractiveness and progress of ten types of NSFC projects when other levels are selected as evaluation contexts. In Level 1, P3 is the most attractive. Instead, P2

$$E^1 = \left\{ \begin{array}{l} \text{National Science Fund for Distinguished Young Scholars (P1), Major Research Plan (P2),} \\ \text{Science Fund for Creative Research Groups (P3)} \end{array} \right\}$$

$$E^2 = \{ \text{Major Program (P4), Joint Funds (P5)} \},$$

$$E^3 = \left\{ \begin{array}{l} \text{Young Scientists Fund (P6), Key Program Projects (P7),} \\ \text{Special Fund for Research on National Major Research Instruments (P8)} \end{array} \right\}$$

$$E^4 = \{ \text{Fund for Less Developed Regions (P9), General Program (P10)} \}.$$

The bracket at the right of the project name denotes the code of the project in this paper.

The efficient score of P1, P2 and P3 is one. They establish the first efficiency frontier. The two projects on the fourth frontier line are P9 and P10, which can, thus, be considered to have the worst scientific producing performance. Using the

shows the minor attractiveness in Level 1. In Levels 2–4, the projects are ranked with reference to attractiveness and progress scores. P10, the General Program, can be considered the worst perform project of highly funded scientists, which is classified into the lowest performance level and has a large room for improvement. Its proportion is 24.07% and ranks

Fig. 1 Results of scientists clustering with the k-means method based on the history of their careers. The “Career Duration” in the figure is equal to the year when this scientist’s last paper was published minus the year when he/she began his/her career. The “Beginning of Career” in the figure is the year when the scientist started to work

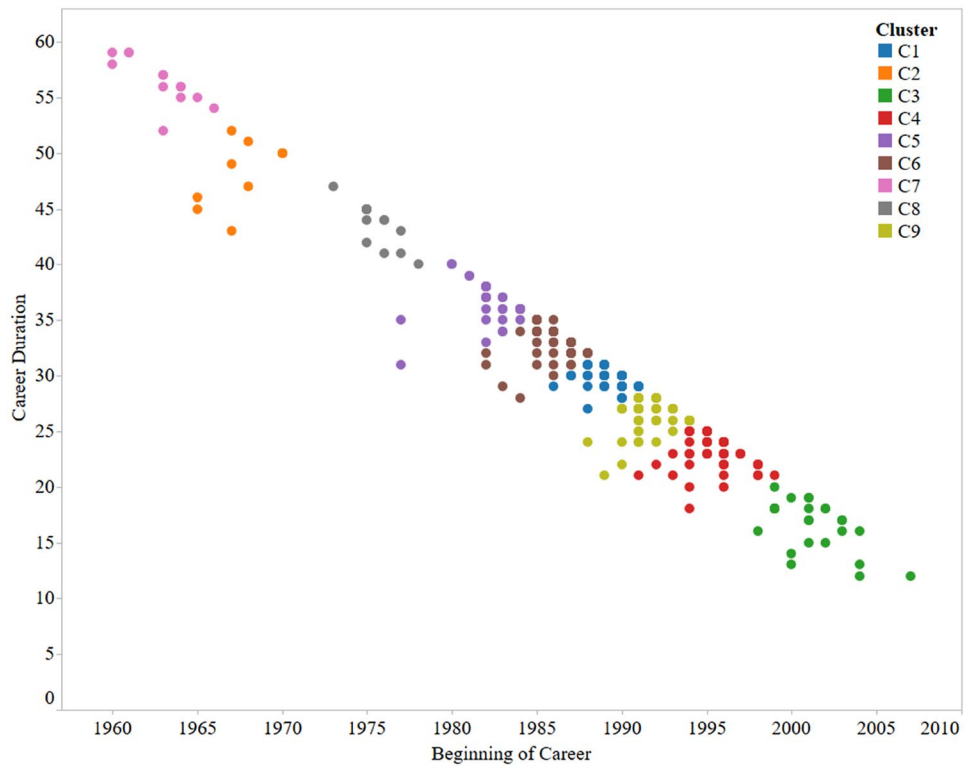


Table 7 Rank of clustering results according to the average beginning year of scientists’ careers from earliest to latest

Cluster	C1	C2	C3	C4	C5	C6	C7	C8	C9
Rank	6	2	9	8	4	5	1	3	7

third of all funding project types. We can surmise that much of the funding waste of highly funded scientists comes from the General Program. Focusing on improving the efficiency of the General Program of highly funded scientists may be helpful to improve their overall R&D efficiency. However, since the General Program, which always represents the small grants, always cares for better states for adventures and has a more considerable influence on research performance [4, 5, 13], relative policies must provide adequate room for the General Program like small grants possible playing a pivotal role in creative research actions. In a nutshell, we should balance the amount of funding across all grant types appropriately. These results may be helpful in management authorities balancing the funding between different types of NSFC funding.

Table 8 Attractiveness and Progress scores for nine clusters of highly funded scientists by context-dependent DEA

	Cluster of scientists	Level 1	Level 2	Level 3
Level 1	C1		1.364(3)	1.506(3)
	C2		1.745(1)	1.926(1)
	C3		1.551(2)	1.732(2)
	C4		1.342(4)	1.481(4)
Level 2	C6	1.203(2)		1.151(2)
	C8	1.074(1)		1.463(1)
Level 3	C5	1.380(1)	1.151(2)	
	C7	2.105(3)	1.457(3)	
	C9	1.479(2)	1.101(1)	

Implication

The progress scores of the NSFC projects denote that the bigger the progress score of a project, the bigger effort the project needs to keep up with the projects which show high efficiency. Similarly, the smaller the progress score of a project, the less effort the project needs to keep up with the projects which show high efficiency. Relative management authorities can select a gradual approach to help improve the performance of highly funded scientists in projects at a low level by adjusting the funding location.

Efficiency analysis of highly funded scientists in different career stages

The career length of highly funded scientists selected spans widely ranging from 12 to 60. To study the relationship between the R&D efficiency and scientists' career stages, we, using the *k*-means method [20], group these scientists into nine clusters according to their career duration and the beginning year of their careers. The result of clustering is depicted in Fig. 1. All clusters are ranked according to their average beginning of the careers from earliest to latest, presented in Table 7, which indicates that cluster C7 started the career earliest and cluster C3 the latest. A smaller rank means a younger group.

Performance level

Using the BCC-DEA production frontier, we grade the nine scientist clusters into 3 levels based on their research efficiency. They are

$$E^1 = \{C1, C2, C3, C4\},$$

$$E^2 = \{C6, C8\},$$

$$E^3 = \{C5, C7, C9\}.$$

The efficient score of C1, C2, C3, and C4 is one. They establish the first efficiency frontier line. The three clusters on the third frontier line are C5, C7, and C9, which can, thus, be considered to have the worst scientific producing performance. The average rank of clusters in the first efficiency frontier line is 6.25, and the second is 4.00, the third is 4.33. It is not hard to find that these average ranks are closer to 5, the middle rank. Highly funded scientists who are in their early career stage may have a slight advantage in R&D efficiency.

Attractiveness and progress

Table 8 presents the attractiveness and progress of nine clusters of highly funded scientists when other levels are selected as evaluation contexts. It is easy to see that the gap is not

big between different levels. We surmise that they already have a relatively stable career when they become a highly funded scientist.

Discussion and conclusion

In this paper, we measured the R&D efficiency of highly funded scientists in China over the timespan covered by our dataset (1986–2019) from multiple perspectives. Using the DEA model, we estimated the R&D efficiency of each scientist. They were appropriately classified into multiple efficiency frontiers. Attractiveness and progress score of every scientist is given, which can help us to deeply understand the R&D efficiency of this scientist in this academic division. Each scientist graded into a less efficient level is suggested a gradual and appropriate learning path to boost their R&D efficiency. Moreover, the R&D efficiency in different types of NSFC projects was analyzed to further discern the deep patterns of highly funded scientists. We found that three sorts of projects were graded into the first tier, including the Major Research Plan, National Science Fund for Distinguished Young Scholars, and Science Fund for Creative Research Groups. In addition, to investigate the impact of career length on the R&D efficiency of highly funded scientists, we first grouped scientists into nine clusters. Then, the R&D efficiency of those clusters was analyzed by the DEA model. It was found that the career length of highly funded scientists had a limited impact on efficiency. We believed that these results would be beneficial for authorities to make policies that had a significant impact on the R&D.

We must point out that our empirical analysis has limitations. First, we only select the three most common indicators for analysis. If we introduce more indicators for analysis, the results may be different and more attractive. Second, we choose highly funded scientists in China for study, who may not fully reflect the characteristics of highly funded researchers worldwide. If we had access to more data about highly funded scientists in other countries, we should get more valid results.

In a word, our main contribution is to introduce and quantitatively analyze the particular group, i.e., the highly funded scientists. We want to call for attention for this specific group. Besides, we exhibit new methodological ideas that focus on showing the capabilities of DEA methods as a grading tool. This tool can grade DMUs into multi-levels. Then, it gives a relatively gradual and appropriate learning path for those DMUs that don't perform well to boost their efficiency without considering improving the accuracy of DEA analysis results.

In fact, many researchers have paid attention to the patterns of particular groups (e.g., Nobel Prize laureates [16–19] and Highly Cited Researchers [6, 7, 23]) to boost

the development of the whole scientific community. Nevertheless, with the scientific funding got increasing attention, these highly funded scientists, which are so unique and significant, have not caught the researchers' appropriate attention. In addition, the collaborative networks of the highly funded scientist's group and the influence of their status on their careers are also exciting and deserve us to study. Another exciting direction for future work is to build models predicting the emergence of highly funded scientists.

Funding This work was supported by the Grants (No. 71871149) from NSF of China, the grants (Nos. sksyl201705, Nos. sksyl201819 and sksyl201821) from Sichuan University, and the grants (No. 18YJC630045) from the MOE (Ministry of Education in China) Project of Humanities and Social Sciences.

Availability of data and materials Due to the size, the raw dataset used is freely available upon request from the corresponding author.

Declarations

Conflict of interests On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Amin GR, Siddiq FK (2019) Measuring global prosperity using data envelopment analysis and OWA operator. *Int J Intell Syst* 34:2713–2738. <https://doi.org/10.1002/int.22176>
- Bagheri M, Ebrahimnejad A, Razavyan S et al (2021) Solving fuzzy multi-objective shortest path problem based on data envelopment analysis approach. *Complex Intell Syst* 7:725–740. <https://doi.org/10.1007/s40747-020-00234-4>
- Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manag Sci* 30:1078–1092. <https://doi.org/10.1287/mnsc.30.9.1078>
- Bloch C, Schneider JW, Sinkjaer T (2016) Size, accumulation and performance for research grants: examining the role of size for centres of excellence. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0147726>
- Bloch C, Sorensen MP (2015) The size of research funding: trends and implications. *Sci Public Policy* 42:30–43. <https://doi.org/10.1093/scipol/scu019>
- Bornmann L, Bauer J (2015) Evaluation of the highly-cited researchers' database for a country: proposals for meaningful analyses on the example of Germany. *Scientometrics* 105:1997–2003. <https://doi.org/10.1007/s11192-015-1619-1>
- Bornmann L, Bauer J, Schlagberger EM (2017) Characteristics of highly cited researchers 2015 in Germany. *Scientometrics* 111:543–545. <https://doi.org/10.1007/s11192-017-2248-7>
- Calik A, Pehlivan NY, Kahraman C (2018) An integrated fuzzy AHP/DEA approach for performance evaluation of territorial units in Turkey. *Technol Econ Dev Econ* 24:1280–1302. <https://doi.org/10.3846/20294913.2016.1230563>
- Castellano C, Radicchi F (2009) On the fairness of using relative indicators for comparing citation performance in different disciplines. *Arch Immunol Ther Exp* 57:85–90. <https://doi.org/10.1007/s00005-009-0014-0>
- Charnes A, Cooper WW, Rhodes E (1978) Measuring efficiency of decision making units. *Eur J Oper Res* 2:429–444. [https://doi.org/10.1016/0377-2217\(78\)90138-8](https://doi.org/10.1016/0377-2217(78)90138-8)
- Dimke H, Norn MT, Christiansen PM et al (2019) Most scientists prefer small and mid-sized research grants. *Nat Hum Behav* 3:765–767. <https://doi.org/10.1038/s41562-019-0643-1>
- Ervural BC, Zaim S, Delen D (2018) A two-stage analytical approach to assess sustainable energy efficiency. *Energy* 164:822–836. <https://doi.org/10.1016/j.energy.2018.08.213>
- Fortin J-M, Currie DJ (2013) Big science vs. little science: how scientific impact scales with funding. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0065263>
- Garfield E, Merton RK (1979) Citation indexing: its theory and application in science, technology, and humanities. Wiley, New York
- Lee H, Choi Y, Seo H (2020) Comparative analysis of the R&D investment performance of Korean local governments. *Technol Forecast Soc Change*. <https://doi.org/10.1016/j.techfore.2020.120073>
- Li J, Yin Y, Fortunato S et al (2019) A dataset of publication records for Nobel laureates. *Sci Data*. <https://doi.org/10.1038/s41597-019-0033-6>
- Li J, Yin Y, Fortunato S et al (2020) Scientific elite revisited: patterns of productivity, collaboration, authorship and impact. *J R Soc Interface*. <https://doi.org/10.1098/rsif.2020.0135>
- Li JC, Yin Y, Fortunato S et al (2019) Nobel laureates are almost the same as us. *Nat Rev Phys* 1:301–303. <https://doi.org/10.1038/s42254-019-0057-z>
- Liang GQ, Hou HY, Ding Y et al (2020) Knowledge recency to the birth of Nobel Prize-winning articles: gender, career stage, and country. *J Informetr* 14:14. <https://doi.org/10.1016/j.joi.2020.101053>
- Macqueen J (1967) Some methods for classification and analysis of multivariate observations
- Moed HF (2010) Measuring contextual citation impact of scientific journals. *J Informetr* 4:265–277. <https://doi.org/10.1016/j.joi.2010.01.002>
- Pakkar MS (2016) Multiple attribute grey relational analysis using DEA and AHP. *Complex Intell Syst* 2:243–250. <https://doi.org/10.1007/s40747-016-0026-4>
- Parker JN, Lortie C, Allesina S (2010) Characterizing a scientific elite: the social characteristics of the most highly cited scientists in environmental science and ecology. *Scientometrics* 85:129–143. <https://doi.org/10.1007/s11192-010-0234-4>
- Qin X, Du D, Kwan M-P (2019) Spatial spillovers and value chain spillovers: evaluating regional R&D efficiency and its spillover effects in China. *Scientometrics* 119:721–747. <https://doi.org/10.1007/s11192-019-03054-7>
- Radicchi F, Fortunato S, Castellano C (2008) Universality of citation distributions: toward an objective measure of scientific

- impact. *Proc Natl Acad Sci USA* 105:17268–17272. <https://doi.org/10.1073/pnas.0806977105>
26. Schubert T, Yang G (2016) Institutional change and the optimal size of universities. *Scientometrics* 108:1129–1153. <https://doi.org/10.1007/s11192-016-2015-1>
27. Seiford LM, Zhu J (2003) Context-dependent data envelopment analysis—measuring attractiveness and progress. *Omega Int J Manag Sci* 31:397–408. [https://doi.org/10.1016/s0305-0483\(03\)00080-x](https://doi.org/10.1016/s0305-0483(03)00080-x)
28. Wang Y (2015) Towards the abstract system theory of system science for cognitive and intelligent systems. *Complex Intell Syst* 1:1–22. <https://doi.org/10.1007/s40747-015-0001-5>
29. Yang G, Ahlgren P, Yang L et al (2016) Using multi-level frontiers in DEA models to grade countries/territories. *J Informetr* 10:238–253. <https://doi.org/10.1016/j.joi.2016.01.008>
30. Zemtsov S, Kotsemir M (2019) An assessment of regional innovation system efficiency in Russia: the application of the DEA approach. *Scientometrics* 120:375–404. <https://doi.org/10.1007/s11192-019-03130-y>
31. Zhang H, Patton D, Kenney M (2013) Building global-class universities: assessing the impact of the 985 Project. *Res Policy* 42:765–775. <https://doi.org/10.1016/j.respol.2012.10.003>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.