Efficiency Evaluation of China's Regional Innovation Process Based on a Network Data Envelopment Analysis Model with a Mixed Structure: Enlightenment from the Regional Tobacco Industry

Zhu Yu, Lecturer
Yang Feng, Professor
Wang Dawei, Ph.D. candidate

Zhu Yu, Ph.D. candidate in Decision-making methods and applications, School of Management, University of Science and Technology of China, Hefei, Anhui 230026, PR China, Lecturer in Economics, School of Economics and Management, Anhui Polytechnic University, Wuhu, Anhui 241000, PR China. Yang Feng, Ph.D. Professor in Operational research, School of Management, University of Science and Technology of China, Hefei, Anhui 230026, PR China, Wang Dawei, Ph.D. candidate in Decision-making methods and applications, School of Management, University of Science and Technology of China, Hefei, Anhui 230026, PR China. Correspondence author: Zhu Yu, Email: zhuyu16@mail.ustc.edu.cn

Abstract: Accurate measurement of regional efficiency is a prerequisite for effective management. Prior studies have expanded on the overall "black box" evaluation with two stages of research and development (R&D) and commercialization, opening up the internal structure of the regional innovation process, but ignoring the independent innovation activities of universities, research institutes, and firms in the R&D stage. We construct a mixed structure with two stages, three actors, and four subsystems, and conduct an empirical analysis of China's provincial samples from 2017 to 2019 by using the network data envelopment analysis (DEA) model. Results show that the efficiency of the R&D stage at the provincial level is generally higher than that of the commercialization stage. However, the three subsystems of the R&D stage perform poorly. Spearman's rank correlation coefficient suggests that there is a significantly positive correlation between total regional efficiency and commercialization. In addition, we use the k-means method to divide 27 provinces into three clusters, setting a more appropriate improvement benchmark for inefficient provinces. Based on enlightenment of regional tobacco industry, we put forward some proposals for specific stage and specific subsystem.

Keywords: Regional innovation efficiency; Network DEA; Mixed structure; Tobacco industry

Tob Regul Sci.™ 2021;7(6): 5184-5196
DOI: doi.org/10.18001/TRS.7.6.11

INTRODUCTION

Innovation is widely recognized to improve national competitiveness. Governments around the world are giving increasing weight to evaluating internal regional efficiency to implement better management.[1] Efficiency here refers to technical efficiency, that is, maximizing innovation outputs while innovation inputs remain unchanged or minimizing innovation inputs while innovation outputs remain unchanged.[2]

In terms of efficiency measurement, there are two popular methods, stochastic frontier analysis (SFA) and data envelopment analysis (DEA). The former is a parametric approach, and the accuracy of the measurement mainly depends on the correctness of the function estimation. The latter is a non-parametric approach that can offer greater flexibility and deal with the issue of multiple inputs and multiple outputs, but assumes that all deviations are due to inefficiency.[2] In general, the two methods have advantages and disadvantages and should be selected according to the specific circumstances. In this paper, we focus on innovation efficiency, which is typically a multi-input and multi-output problem. Owing to the diversity of actors and the complexity of the production process, DEA is a more appropriate approach.

In recent years, scholars have employed the DEA method to calculate innovation efficiency in the context of different countries.[1-4] However, these studies only considered the overall inputs and outputs of the region using the traditional DEA model, making it easy to overestimate efficiency and fail to provide managers with more information.[2] Some researchers refered to the innovation process as the stage of R&D and commercialization.[5] The former is the stage of transforming innovation inputs into knowledge products, and the latter is the stage of manufacturing the created knowledge into products that are brought to the market. The output of the former affects the input of the latter, affecting the overall efficiency; that is, R&D and commercialization are related.

Universities, research institutes, and firms are key actors in the two-stage process. There are three actors develop technology and create knowledge in the R&D stage.[6] In the commercialization stage, firms transform knowledge into products and introduce them to the market.[5] we need to further analyze the efficiency levels of the three actors to implement management that is more targeted. To the best of our knowledge, there are no publications focusing on this key point. To explore the issue, we suppose that regions form a mixed-structure network in two stages, three actors, and four subsystems, and proposed a mixed network DEA model.

Since the late 1990s, Chinese policy makers have invested ample technological resources to enhance the level of innovation.[7] R&D expenditure and R&D personnel in 2018 reached RMB 1.96 trillion and 4.38 million people, which increased by 29-fold and 5-fold since 1998, respectively.[8] The number of patent applications and grants saw 33-fold and 26-fold growth, respectively (up from 1999 to 2019).[9] Although these achievements were impressive, there are still criticisms of the low patent conversion rate and the innovation inefficiency.[7] In 2020, China still ranked 14th in the Global Innovation Index.[10] In this context, our evaluation of China's regional innovation efficiency is of great significance.

Based on enlightenment of regional tobacco industry, we find that, compared with the commercialization stage, the R&D stage has higher efficiency. Also, there is an obvious difference among the three actors in the R&D stage. Furthermore, to find a suitable reference benchmark for the inefficient provinces, we adopted the k-means clustering method to divide 27 provinces into three clusters, and the province in each cluster that performed well can be used as a target for imitation and catch-up by other provinces in the cluster.

The major contributions are as follows. As far as we know, this is the first time that the innovation activities of the three agents of universities, research institutes, and firms have been considered in two stages at the same time and combined with the network DEA model for evaluation. Previous literature used the two-stage DEA to obtain only R&D and commercialization stage efficiency, which can help formulate innovation policies for a specific subsystem, but we expand upon these ideas. Given that China has included the improvement of innovation efficiency in the outline of its national strategy program,[11] this paper can be of academic and practical value for researchers and managers who are interested in China's regional innovation. Broadly speaking, given the prevalent regional innovation efficiency evaluation in other countries, the mixed structure we propose can provide necessary knowledge for them.

This study is list as follows. We lay out the literature review in section 2. In section 3, we introduce the model. The empirical study is described in section 4. We offers a discussion in section 5. In section 6, we presents the conclusions.

LITERATURE REVIEW

Network DEA

Classical DEA models, The CCR and BCC models, as classical DEA model, do not explore the internal structures of decision-making units (DMUs). This shortcoming has spurred some interesting studies aimed at better evaluating DMUs with network structures.

Seiford and Zhu (1999) presented an independent two-stage DEA model.[12]A relational DEA model, where the same factors have the same weight, was developed by Kao and Hwang (2008).[13] The two-stage model is easily generalized to a multiple-stage model, and they are both called series structure models; that is, the intermediate products produced in the former stage can be used in the latter stage.[14]

Some scholars have found that certain production or operation activities are carried out simultaneously in a parallel manner. One of the earliest parallel system studies was by Färe et al. (1997), who investigated the farms' efficiency in Illinois, USA.[15] Yang, Jiang, and Ang (2019) treated a company with multiple branches as a parallel systems and developed a new DEA model.[16]

The abovementioned series and parallel structures are both basic network structures. Combining them can yield a more complex network structure, that is, a mixed network structure. Ang, Liu, and Yang (2020) developed a resource allocation model with multiple parallel two-stage structures, and applied it to 17 city bank branches.[17]

The studies' findings indicate that for DMUs with complex internal structures, the efficiency calculated by the network DEA is more accurate and provides more useful information, such as inefficiency nodes for managers.

Regional innovation efficiency evaluation with DEA

To understand the level of innovation, the measurement of innovation efficiency is a key point[18].Belgin (2019) analyzed the innovation efficiency of 12 Turkish regions though two inputs and two outputs.[4] Zemtsov and Kotsemir (2019) explored the efficiency of regional technology creation in Russia from 1998 to 2012 with DEA.[5]

These studies regarded innovative production as a "black box" and failed to consider the process of innovative production, which may have caused some deficiencies to be ignored. Some studies have begun to focus on decomposing system efficiency from a two-stage perspective. Chen and Guan (2012) applied two-stage DEA to calculate the innovation performance of 30 regions in China.[19] Min, Kim, and Sawng (2020) compared the efficiency deviations in the R&D and commercialization stage in South Korea.[18]

These studies stressed that the transfer of new knowledge created in the first stage extended to the next stage, and finally produced new products. Actually, universities, research institutes, and firms are major innovation agents and directly create new knowledge and technological in R&D.[1] We need to further understand their level of efficiency in the regional innovation system. Zuo and Guan (2017) and Xiong, Yang, and Zhong (2020) were among the few researchers to consider the three agents simultaneously in the R&D process. Nevertheless, these two studies only discussed the R&D stage.[6, 20]

Therefore, the regional innovation process is a complex structure with multiple actors and multiple stages. The network DEA method provides us with valuable evaluation insights, especially the mixed-structure DEA model. Next, we propose a mixed DEA model with two stages and four subsystems.

MIXED STRUCTURE WITH TWO STAGES AND FOUR SUBSYSTEMS

Suppose a production system contains four subsystems, as shown in Fig. 1. Subsystems 1, 2, and 3 all require certain inputs, X_1 and X_2, to organize production. The output, Z_1, produced by subsystem 1, is used as input of subsystem 4, and the other output, Y_1, is used as the output of the whole system. Similarly, subsystem 2 also produces Z_1 and Y_1 as the input and output for subsystem 4 and the whole system, respectively. Additionally, subsystem 3 produces Z_2 as input for system 4. Subsystem 4 utilizes Z_1 and Z_2, produced by subsystem 1, 2 and 3, and the input of the whole system, X_3, to produce Y_2.

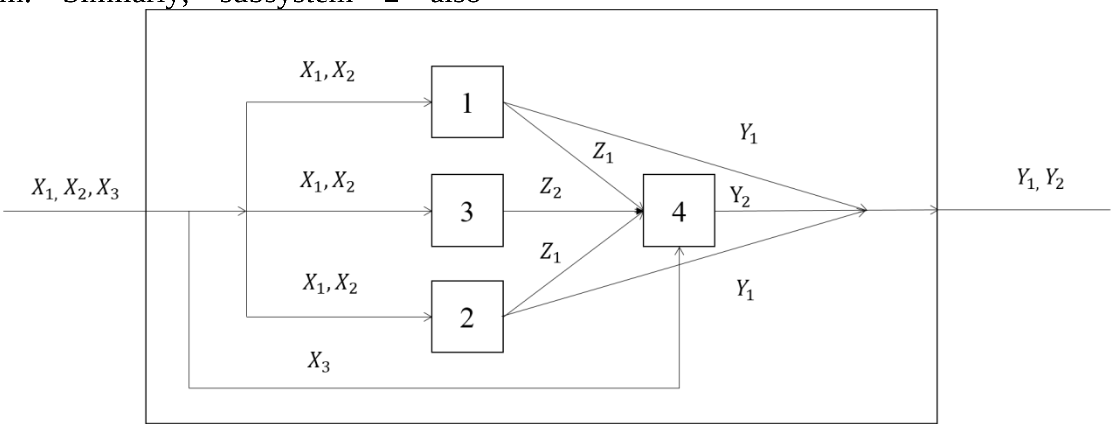

Figure 1 Network system with four subsystems

If we use the traditional CCR model to measure efficiency of the DMUs with network structure as shown in Fig. 1, as we can only see the input X_1, X_2, and X_3 of the overall system, and the final output, Y_1 and Y_2, the complex structure inside the system is ignored. This will cause two problems. First, the whole system's efficiency will be overstated.[14] Second, when the whole system is inefficient, we will not know the reason and improve it. Therefore, opening the "black box" and understanding its component processes are necessary.

Figure 1 shows that subsystems 1 –3 constitute a parallel structure, and they are in a sequential relationship with subsystem 4. To better describe this complex network structure and facilitate modeling, we add two virtual subsystems, 5 and 6, as auxiliary units. The system in Fig. 1 is equivalently denoted as a two-stage series structure (see Fig. 2). The first stage is a parallel system composed of subsystems 1, 2, and 3, and virtual subsystem 5. The second stage is a parallel system as well, which is composed of subsystem 4 and virtual subsystem 6.

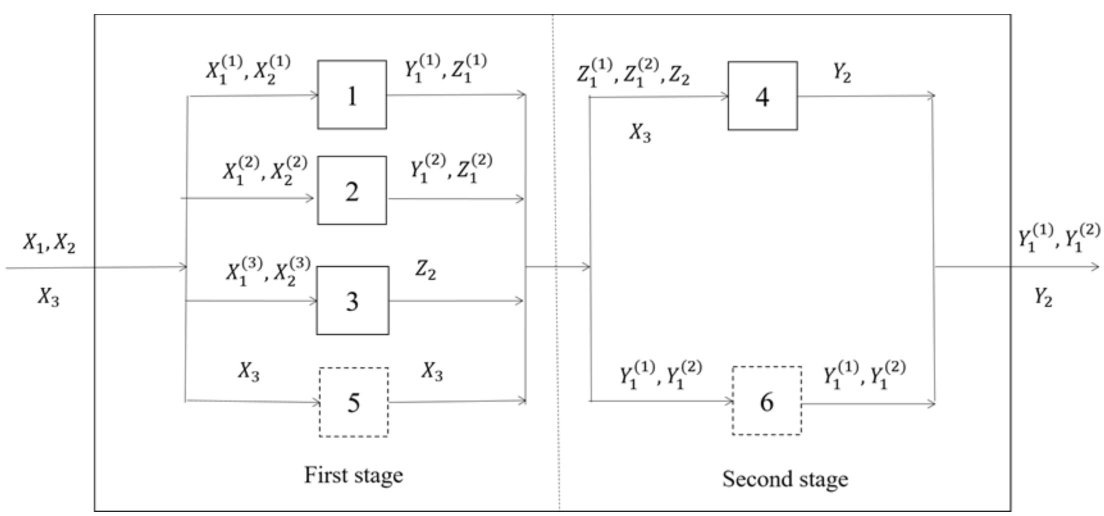

Figure 2 An equivalent mixed structure

Assume there are *n* DMUs, and each $DMU_j (j = 1, \ldots, n)$ has a two-stage production system (see Fig. 2). For DMU_j, $X_{ij} (i = 1, 2, 3)$ denotes the inputs of the first stage. Let $X_{ij}^{(1)}, X_{ij}^{(2)}\ and\ X_{ij}^{(3)}$ denote the *i*th input of subsystem 1, 2, and 3; clearly, we obtain $X_{1j}^{(1)} + X_{1j}^{(2)} + X_{1j}^{(3)} = X_{1j}$, $X_{2j}^{(1)} + X_{2j}^{(2)} + X_{2j}^{(3)} = X_{2j}$. Subsystem 1 produces two kinds of outputs, that is, $Z_{1j}^{(1)}\ and\ Y_{1j}^{(1)}$. $Z_{1j}^{(1)}$ is used by subsystem 4 as an input, and $Y_{1j}^{(1)}$ is the ultimate output. Similarly, the output of subsystem 2 is separated into $Z_{1j}^{(2)}\ and\ Y_{1j}^{(2)}$, where the former is utilized by subsystem 4 for production, and the latter is the final output. Subsystem 3 has just one type of output, Z_{2j}. Like $Z_{1j}^{(1)}\ and\ Z_{1j}^{(2)}, Z_{2j}$ is also used for subsystem 4. Subsystem 4 uses $Z_{1j}^{(1)}, Z_{1j}^{(2)}, Z_{2j}$, and X_{3j} to produce Y_{2j}. The inputs of virtual subsystem 5 are the same as its outputs, as well as virtual subsystem 6. Then, the overall efficiency of the *d*th DMU is:

$$
\begin{aligned}
E_d = \ & \max\ u_1 Y_{1d}^{(1)} + u_1 Y_{1d}^{(2)} + u_2 Y_{2d} \\
s.t.\ & v_1 X_{1d} + v_2 X_{2d} + v_3 X_{3d} = 1 \\
& \left(u_1 Y_{1j}^{(1)} + u_1 Y_{1j}^{(2)} + u_2 Y_{2j}\right) - \left(v_1 X_{1j} + v_2 X_{2j} + v_3 X_{3j}\right) \le 0 \\
& \left(u_1 Y_{1j}^{(1)} + w_1 Z_{1j}^{(1)}\right) - \left(v_1 X_{1j}^{(1)} + v_2 X_{2j}^{(1)}\right) \le 0 \\
& \left(u_1 Y_{1j}^{(2)} + w_1 Z_{1j}^{(2)}\right) - \left(v_1 X_{1j}^{(2)} + v_2 X_{2j}^{(2)}\right) \le 0 \\
& w_2 Z_{2j} - \left(v_1 X_{1j}^{(3)} + v_2 X_{2j}^{(3)}\right) \le 0 \\
& u_2 Y_{2j} - \left(w_1 Z_{1j}^{(1)} + w_1 Z_{1j}^{(2)} + w_2 Z_{2j} + v_3 X_{3j}\right) \le 0 \\
& u_1, u_2, v_1, v_2, v_3, w_1, w_2 \ge \varepsilon \qquad j = 1, \ldots, n
\end{aligned}
\tag{1}
$$

In model (1), $u_1\ and\ u_2$ are the weights of $Y_{1d}\ and\ Y_{2d}$, respectively; $v_1, v_2, and\ v_3$ are the weights of $X_{1d}, X_{2d}, and\ X_{3d}$, respectively; $and\ w_1\ and\ w_2$ are the weights of $Z_{1d}\ and\ Z_{2d}$, respectively. It is worth noting that same indicator has the uniform weight.[5, 14] For instance, for X_1, the weights of $X_{1d}^{(1)}, X_{1d}^{(2)}, and\ X_{1d}^{(3)}$ are all v_1, and so on. The second constraint corresponds to the overall system, and the third to sixth constraint corresponding to subsystems 1 to 4. In addition, owing to the inputs of auxiliary subsystems 5 and 6 being the same as their outputs, the constraint relating to subsystems 5 and 6 is $v_3 X_3 - v_3 X_3 \le 0, \left(Y_1^{(1)} + Y_1^{(2)}\right) - \left(Y_1^{(1)} + Y_1^{(2)}\right) \le 0$, respectively, which always holds.[21] Hence, we do not add the two constraints in model (1). After calculating model (1), we can obtain the optimal weights $u_1^*, u_2^*, v_1^*, v_2^*, v_3^*, w_1^*, and\ w_2^*$. Next, we take the optimal weights to solve the efficiencies of the four subsystems, as follows:

$$E_d^{(1)} = \frac{\left(u_1^* Y_{1d}^{(1)} + w_1^* Z_{1d}^{(1)}\right)}{\left(v_1^* X_{1d}^{(1)} + v_2^* X_{2d}^{(1)}\right)} \tag{2}$$

$$E_d^{(2)} = \frac{\left(u_1^* Y_{1d}^{(2)} + w_1^* Z_{1d}^{(2)}\right)}{\left(v_1^* X_{1d}^{(2)} + v_2^* X_{2d}^{(2)}\right)} \tag{3}$$

$$E_d^{(3)} = \frac{w_2^* Z_{2d}}{\left(v_1^* X_{1d}^{(3)} + v_2^* X_{2d}^{(3)}\right)} \tag{4}$$

$$E_d^{(4)} = \frac{u_2^* Y_{2d}}{\left(w_1^* Z_{1d}^{(1)} + w_1^* Z_{1d}^{(2)} + w_2^* Z_{2d} + v_3^* X_{3d}\right)} \tag{5}$$

Following Kao (2009),[21] we have $E_d = E_d^I \times E_d^{\Pi}$, where $E_d^I\ and\ E_d^{\Pi}$ are reprensted the efficiency value of the first and second stage.

$$E_d^I = \frac{\left(u_1^* Y_{1d}^{(1)} + w_1^* Z_{1d}^{(1)}\right) + \left(u_1^* Y_{1d}^{(2)} + w_1^* Z_{1d}^{(2)}\right) + w_2^* Z_2 + v_3^* X}{v_1^* X_{1d} + v_2^* X_{2d} + v_3^* X_{3d}} \tag{6}$$

$$E_d^{\Pi} = \frac{u_1^* Y_{1d}^{(1)} + u_1^* Y_{1d}^{(2)} + u_2^* Y_{2d}}{\left(w_1^* Z_{1d}^{(1)} + w_1^* Z_{1d}^{(2)} + w_2^* Z_{2d}\right) + \left(u_1^* Y_{1d}^{(1)} + u_1^* Y_{1d}^{(2)}\right)} \tag{7}$$

AN APPLICATION TO PROVINCIAL REGIONS IN CHINA

We use an administrative provincial-level region as the unit of evaluation. The main reasons are as follows: First, China's provinces have clear geographical and administrative boundaries. They have the autonomy to formulate local economic and social policies under the control of the central government.[1,5] Second, during the long-term historical development of China, each province has formed its own unique natural resources, cultural characteristics, and industrial environment. Third, the National Bureau of Statistics of China usually uses provincial-level administrative regions as the basic unit for statistics, and provides data for quantitative analysis at the provincial level. Based on these considerations, many researchers employed the provincial level as the unit of measurement.[2, 19]

Indicators and data

Assume that each province has a regional innovation process, as shown in Fig. 3.

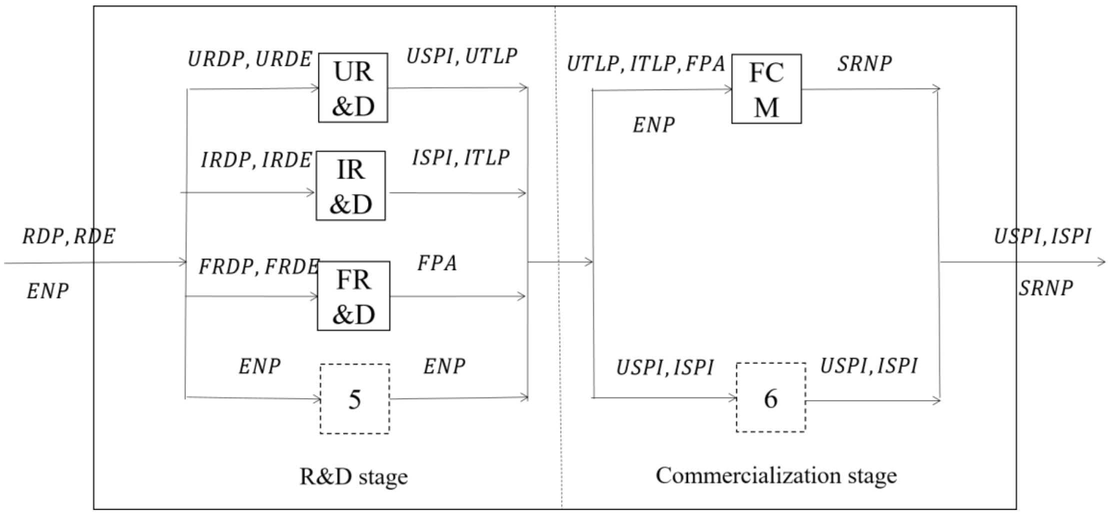

Figure 3 Regional innovation process

The first stage is the R&D stage, which is composed of three R&D subsystems by universities, research institutes, and firms. UR&D, IR&D, and F&D are denoted the three subsystems. The second stage is the commercialization stage, with only one subsystem. The subsystem is denoted by FCM, which refers to the subsystem that transforms knowledge into business. R&D is inseparable from the labor and R&D funds. Therefore, this study uses the full-time equivalent of R&D personnel (RDP) and internal expenditures on R&D (RDE) of each innovation entity as the input of the three subsystems in the first stage.[2, 6, 20] Specifically, URDP and URDE are expressed as the UR&D stage's inputs, IRDP and IRDE are the inputs of IR&D, and the inputs of FR&D are FRDP and FRDE (See Table 1).

Universities and research institutes conduct technological and creation activities in the first stage. Certain research and development results of universities and research institutes will be used for commercial applications. The number of transfers and licenses of patent ownership (TIP) and the scientific papers issued (SPI) are used as indicators to measure outputs, where the former is the intermediate output, and latter is the final output.[22] Similarly, USPI and UTLP are the outputs of UR&D, and the IR&D's outputs are ISPI and ITLP (Table 1).

Unlike the case with universities and research institutes, for most companies, publishing research articles is not helpful for profit and may cause commercial secrets to be leaked.[23] Therefore, the number of firms' patent applications (FPA) is regarded as the R&D output of the firm.[2] Concerning the commercialization stage, expenditure on new products development (ENP) is considered an input, and the sales revenue of new products is regarded as an output indicator.[5] Table 1 presents indicators description. Table 2 displays a descriptive statistics analysis of all variables.

Table 1
Indicators description

Type	Variable	Definition	Unit
Inputs	$X_1^{(1)}$	Full-time Equivalent of R&D Personnel in University (URDP)	Man-year
	$X_2^{(1)}$	Intramural Expenditures on R&D in University (URDE)	10,000 yuan
	$X_1^{(2)}$	Full-time Equivalent of R&D Personnel in Research Institute (IRDP)	Man-year
	$X_2^{(2)}$	Intramural Expenditures on R&D in Research Institute (IRDE)	10,000 yuan
	$X_1^{(3)}$	Full-time Equivalent of R&D Personnel in Firm (FRDP)	Man-year
	$X_2^{(3)}$	Intramural Expenditures on R&D in Firm (FRDE)	10,000 yuan
	X_3	Expenditure on New Products Development (ENP)	10,000 yuan
Intermediate outputs	$Z_1^{(1)}$	Number of Transfers and Licenses of Patent Ownership in University (UTLP)	Piece
	$Z_1^{(2)}$	Number of Transfers and Licenses of Patent Ownership in Research Institute (ITLP)	Piece
	Z_2	Patent Applications of Firm (FPA)	Piece
Final outputs	$Y_1^{(1)}$	Scientific Papers Issued of University (USPI)	Piece
	$Y_1^{(2)}$	Scientific Papers Issued of Research Institute (ISPI)	Piece
	Y_2	Sales Revenue of New Products (SRNP)	10,000 yuan

Table 2
Descriptive statistics for variables

Variable	N	Min	Max	Std. Dev	Mean
$X_1^{(1)}$	27	1068	35207	8166.5	13920.2
$X_2^{(1)}$	27	18188	1828063	435106.3	464078.4
$X_1^{(2)}$	27	1956	102538	19673.5	14822.9
$X_2^{(2)}$	27	97007	7412398	1499438	896008.9
$X_1^{(3)}$	27	1971	457342	124744.2	100802.2
$X_2^{(3)}$	27	74815	18650313	5209891.8	4420428.1
X_3	27	147630	33366963	7558746.6	5523479.3
$Z_1^{(1)}$	27	2	1697	334.9	231.8
$Z_1^{(2)}$	27	2	745	149.6	72
Z_2	27	576	241700	54189.3	35208.4
$Y_1^{(1)}$	27	5354	131706	33950.5	50655.9
$Y_1^{(2)}$	27	1446	58671	10760.4	6360.4
Y_2	27	1053103	393760563	93405756.9	72606367

Moreover, since the conversion of inputs to outputs takes time, we set a one-year lag following Chen et al. (2018).[5] To be specific, the data for RDP and RDE are from 2017, the data for ENP, TLP, FPA, and SPI are from 2018, and the data for SRNP are from 2019. All data are derived from the *China Statistical Yearbook on Science and*

Technology (CSYST). We deflate all of the currency data to the year 2010 using China's Consumer Price Index. Additionally, given the lack of data on TLP in some provinces, after excluding these provinces, we have 27 provinces in mainland China.

Efficiency results

We employ the models proposed in section 3 to calculate the efficiency. $E, E^{I}, \text{ and } E^{\Pi}$ are denoted as overall efficiency and two sub-stages' efficiency. The efficiency of four subsystems are represented by $E^{(1)}$, $E^{(2)}$, $E^{(3)}$, and $E^{(4)}$ (Table 3). The three key results are given below.

First, the overall efficiency of no provinces is effective. Some provinces perform well for the whole system, such as Zhejiang (0.9997) and Tianjin (0.9825), which are about three times as efficient as Gansu (0.3283) and Hainan (0.3887).

Second, the inter-provincial difference in the first stage is much smaller than in the second stage, and the average efficiency of the former is higher than that of the latter. In the first stage, many provinces are close to effective, and the lowest is Jiangxi (0.8334). There is one efficient province Zhejiang in the second stage, while the lowest province is Gansu (0.3283).

Third, among the four subsystems, the highest average efficiency is that of FCM, and the lowest is IR&D. The UR&D in Jiangsu and Inner Mongolia are effective, the FR&D in Anhui and Jiangxi are effective, and the FCM in Zhejiang is also effective. As the agent of the two subsystems, FCM and FR&D, firms have inconsistent efficiency.

Table 3
Innovation efficiencies in 27 provinces of China

Province	E	E^{I}	E^{Π}	$E^{(1)}$	$E^{(2)}$	$E^{(3)}$	$E^{(4)}$
East							
Beijing	0.6251	0.9988	0.6259	0.3473	0.0659	0.5939	0.6257
Tianjin	0.9825	0.9996	0.9829	0.2414	0.0362	0.4808	0.9827
Hebei	0.7373	0.9996	0.7376	0.7129	0.0256	0.3672	0.7375
Shanghai	0.7060	0.9990	0.7067	0.3908	0.0523	0.4205	0.7065
Jiangsu	0.6273	0.9978	0.6286	1.0000	0.0443	0.6930	0.6277
Zhejiang	0.9997	0.9988	1.0000	0.7378	0.1833	0.7437	1.0000
Fujian	0.5815	0.9995	0.5818	0.3441	0.1091	0.5415	0.5817
Shandong	0.7150	0.9983	0.7162	0.5628	0.1136	0.3028	0.7159
Guangdong	0.6471	0.8419	0.7686	0.5090	0.0943	0.9798	0.6535
Hainan	0.3887	0.9999	0.3887	0.9911	0.0963	0.5571	0.3887
Center							
Shanxi	0.7810	0.9999	0.7811	0.8598	0.0774	0.3707	0.7811
Anhui	0.9093	0.9995	0.9098	0.6790	0.0475	1.0000	0.9094
Jiangxi	0.6719	0.8334	0.8062	0.7016	0.0711	1.0000	0.6805
Henan	0.9346	0.9995	0.9352	0.7853	0.0365	0.4493	0.9349
Hubei	0.8248	0.9994	0.8253	0.5904	0.0511	0.4618	0.8251
Hunan	0.7659	0.9995	0.7664	0.8221	0.0346	0.4410	0.7662
West							
Inner Mongolia	0.7888	0.9999	0.7889	1.0000	0.0452	0.2689	0.7888
Guangxi	0.9157	0.9999	0.9158	0.5222	0.1338	0.5169	0.9158
Chongqing	0.7424	0.9997	0.7426	0.4711	0.1197	0.4985	0.7425
Sichuan	0.4955	0.9994	0.4958	0.6363	0.0211	0.6724	0.4957
Guizhou	0.5528	0.9999	0.5529	0.4827	0.0852	0.7064	0.5528
Yunnan	0.5242	0.9999	0.5243	0.7366	0.0709	0.5384	0.5242
Shaanxi	0.4712	0.9996	0.4714	0.9379	0.0208	0.3999	0.4714

Gansu	0.3283	0.9999	0.3283	0.9356	0.0706	0.5533	0.3283
Northeast							
Liaoning	0.8106	0.9996	0.8110	0.5058	0.0254	0.3519	0.8109
Jilin	0.6806	0.9999	0.6807	0.7255	0.0907	0.3410	0.6807
Heilongjiang	0.5739	0.9999	0.5740	0.5122	0.0576	0.2566	0.5739
Mean	0.6956	0.9875	0.7054	0.6571	0.0696	0.5373	0.6964
Std.Dev	0.1744	0.0432	0.1758	0.2159	0.0391	0.2091	0.1744

To further explore the relationship among the total system efficiency, substage efficiencies and subsystem efficiencies, we conduct the Spearman's rank correlation coefficient (see Table 4). Table 4 shows that the efficiency value of the FCM and commercialization stage is significantly positively correlated with the whole region. The FCM and the commercialization stage are also significantly positively correlated. Furthermore, none of the other correlation coefficients pass the significance test.

Table 4
Spearman's rank correlation coefficient results

	E	E^{I}	E^{II}	$E^{(1)}$	$E^{(2)}$	$E^{(3)}$	$E^{(4)}$
E	1						
E^{I}	−0.211	1					
E^{II}	0.962**	−0.322	1				
$E^{(1)}$	−0.125	0.205	−0.142	1			
$E^{(2)}$	−0.002	0.104	0.031	−0.190	1		
$E^{(3)}$	−0.165	−0.322	−0.003	−0.079	0.251	1	
$E^{(4)}$	1.000**	−0.211	0.962**	−0.125	−0.002	−0.165	1

Cluster analysis

To explore China's innovation efficiency from a broader viewpoint, some scholars have divided provinces into four geographic areas (east, center, west, and northeast, see Table 3) and analyzed.[5, 20] By observing the efficiency results, we find that, in Jiangsu and Tianjin, which are both eastern provinces, the UR&D efficiency of the former is effective, and that of the latter is only 0.2414. In addition, Anhui and Henan, in the central region, and Guangxi, in the western region, also rank in the forefront in overall efficiency. This indicates that classifying by geography does not adequately reflect the similarity of the innovation performance of each province.

Next, we adopt the K-means technique, a widely used clustering method,[24,25] to classify 27 provinces according to the overall efficiencies and the efficiencies of substages and subsystems, as calculated in Section 4.2 (using SPSS software). Table 6 gives the results of 27 provinces divided into three classifications. The seven mean efficiencies of three classifications are presented in Fig. 4.

For the cluster 1, the whole system (E), the commercialization stage (E^{II}), the IR&D $\left(E^{(2)}\right)$, and FCM $\left(E^{(4)}\right)$ are the least efficient, and the R&D stage (E^{I}) and UR&D $\left(E^{(1)}\right)$ are the most efficient. Cluster 2 is the polar opposite of cluster 1. All the efficiency averages of cluster 3 are in the middle of the three clusters, except for FR&D $\left(E^{(3)}\right)$. Additionally, cluster 3 has the most provinces.

Table 5
Clustering result of 27 provinces

Classification	Provinces
Cluster 1	Hainan, Jiangsu, Sichuan, Yunnan, Guizhou, Gansu, Shaanxi
Cluster 2	Tianjin, Zhejiang, Anhui, Jiangxi, Guangdong, Guangxi
Cluster 3	Beijing, Shanghai, Hebei, Hunan, Inner Mongolia, Liaoning, Shanxi, Jilin, Henan, Heilongjiang, Fujian, Shandong, Hubei, Chongqing

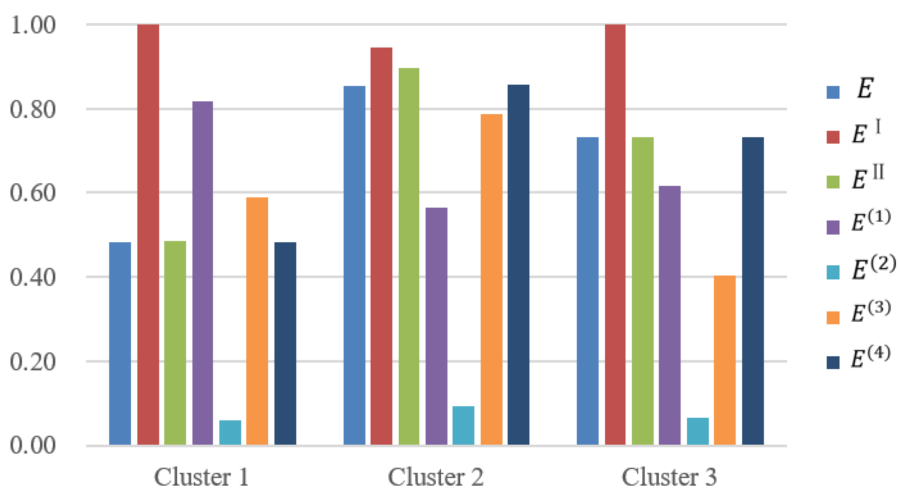

Figure 4 Clustering result of 27 provinces

DISCUSSION

This study propose a measurement framework for the regional innovation process to find inefficient nodes, so that policy makers and researchers can better understand regional innovation activities and efficiency levels.

First, our empirical results suggest that R&D efficiency is generally high, and the commercialization efficiency score is relatively low. The results are different from previous studies. For example, Chen and Guan (2012) found that most provinces in China perform inefficiently, and the average efficiency in the commercialization stage is higher than that in the R&D stage.[19] Min et al. (2020) indicated that the average efficiency of the two stages in Korea is basically balanced over five time points.[18] As the DEA approach measures the relative efficiency, these mixed findings may be caused by different data sets or models.

Moreover, there is a significant positive correlation between the efficiency of the whole system and the efficiency of commercialization stage. Therefore, the inefficiency in many provinces is accompanied by inefficiency in the commercialization stage. This suggests that substantial technical achievements have encountered obstacles in the process of bringing them to the market. Hence, the pursuit of higher R&D investment and more patent output is not the driving force to improve innovation efficiency. Thus, it is important to formulate policies that incentivize achievements and transformation.

Second, this paper goes beyond prior research that ignored the activities of the three agents in the R&D stage. Specifically, if we only look at the efficiency of the first stage, almost all provinces are close to effective, but the three subsystems of the first stage perform poorly, especially the research institutes. Research institutes invest more than universities in R&D, but produce less (see Table 2). Moreover, although the actors in the two subsystems of FR&D and FCM are firms, the national average efficiency of FR&D is significantly lower than that of FCM (See Table 3), which indicates that the firms are better at marketization than R&D.

Third, in 2018, the number of patent applications by universities and research institutes were 320,790 and 61,404, respectively, while the UTLP and ITLP were 6,265 and 1,959 (CSYST, 2019). Obviously, a large number of R&D achievements have become sleeping technologies, which are neither transferred nor licensed to enterprises. A reasonable explanation is that the number of published papers and patent applications is the most important indicator for Chinese scientists, and they are unwilling to spend limited energy on patent conversion.[7]

Fourth, unlike some research conducted on geographic classification,[5, 20] we divide areas into three classifications according to the overall, subsystem, and substage efficiencies of each province. The classification for provinces can help poor performers set a more suitable benchmark for improving their efficiencies. For instance, for Gansu and Hainan, it is unrealistic to expect Zhejiang to improve its goals in the beginning. They should take Jiangsu, the best performer in cluster 1, as an improvement target.

Policy implications

This research provides some insights for policy makers and managers.

First, owing to the interrelationship between two stages, they should be aware of conflicting pressure on the intermediate output in the two stages; that is, the R&D stage requires more intermediate output to improve performance, while the second stage needs the intermediate output to be as small as possible. This should remind managers that the policy of promoting patent conversion and output should be accompanied by a supporting policy to improve the level of commercialization.

Second, promoting patent transformation and improving commercialization efficiency may be related to the following point. As the current evaluation standards still focus more on the number of publications and patents rather than applications quality,[7] it is necessary to reform the evaluation mechanism of universities and research institutes, combine the characteristics of disciplines, strengthen the proportion of patented technology applications in the evaluation system, and reform the distribution of benefits approaches to encourage researchers to engage in technological transformation. However, the government should also pay attention to cultivate talent with both professional technology and marketing skills and establish more platforms for patent transfer and identification so as to promote collaboration communication among universities, research institutions, and firms.

Third, owing to differences between provinces, policies should be formulated for stage-specific and subsystem-specific scenarios. From the perspective of the two sub-stages, the provinces in cluster 1 should further improve the efficiency of commercialization and transform more knowledge achievements into economic value. The regions in cluster 2 should focus on new technology development. Furthermore, the regions in cluster 3 could be considered to enhance the performance of the two stages at the same time. From the perspective of subsystems, for all provinces, there is a big margin for improving the research institutes' efficiency. By improving efficiency and reasonably reducing R&D investment, the output performance of research institutes may be increased. For some provinces with low efficiency among universities (e.g., Tianjin), they can pay more attention to the real demand of the market. Firms in some provinces need to improve R&D efficiency first, such as Heilongjiang and Inner Mongolia, while firms in some provinces need to improve commercialization efficiency, such as Hainan and Gansu. There are also firms in several provinces (e.g., Shaanxi) whose efficiency in knowledge R&D and marketization needs to be improved simultaneously. Firms should realize that R&D is the source of power to ensure product innovation and should actively cooperate with the other actors, that is, universities and research institutes.

CONCLUSIONS

We analyzed the innovation structure of China's provinces and assessed them by using the mixed network DEA . The constructed hybrid structure reveals weaknesses within each province. We find that the efficiency score of R&D stage is generally higher, while the total efficiency of each province is mainly subject to the commercialization efficiency. There is a large difference regarding the efficiency of the different subsystems. The IR&D has the lowest average efficiency, and the other three subsystems have significant regional differences, so they should implement targeted policies. Furthermore, we use k-means to divide all provinces into three clusters. The province with the best performance in each cluster can be used as a benchmark for the improvement of other provinces in the cluster.

Although we propose a hybrid model for evaluating regional innovation efficiency, there are still some future research directions. First, we can consider the interaction between consecutive periods, that is, dynamic efficiency evaluation. Second, because universities, research institutes, and firms have different organizational goals, adding their preferences and characteristics to the model may produce some interesting results. Third, following Chen et al. (2018),[5] this paper attempts to find the inefficient nodes by exploring the internal process of regional innovation. However, the specific influencing factors and solutions require more research.

CONFLICTS OF INTEREST DISCLOSURE STATEMENT

This research is not funded by any organization related to tobacco production.

ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundation of China (Grant Nos. 71631006, 71991464, and 71921001).

REFERENCES

1. Li, Xibao. "China's regional innovation capacity in transition: An empirical approach." *Research policy*. 2009; 38(2): 338-357. doi:https://doi.org/10.1016/j.respol.2008.12.002

2. Chen, Kaihua, Mingting Kou, and Xiaolan Fu. "Evaluation of multi-period regional R&D efficiency: An application of dynamic DEA to China's regional R&D systems." *Omega*. 2018; (74): 103-114. doi: https://doi.org/10.1016/j.omega.2017.01.010

3. Belgin, O. Analysing R&D efficiency of Turkish regions using data envelopment analysis. *Technology Analysis & Strategic Management*. 2019; 31(11): 1341-1352. doi: https://doi.org/10.1080/09537325.2019.1613521

4. Zemtsov, S., & Kotsemir, M. An assessment of regional innovation system efficiency in Russia: the application of the DEA approach. *Scientometrics*. 2019; *120*(2): 375-404. doi:https://doi.org/10.1007/s11192-019-03130-y

5. Chen, X., Liu, Z., & Zhu, Q. Performance evaluation of China's high-tech innovation process: Analysis based on the innovation value chain. Technovation. 2018; 74: 42-53. doi:https://doi.org/10.1016/j.technovation.2018.02.009

6. Zuo, K., & Guan, J. Measuring the R&D efficiency of regions by a parallel DEA game model. *Scientometrics*. 2017; 112(1): 175-194. doi:https://doi.org/10.1007/s11192-017-2380-4

7. Liu, X., Schwaag Serger, S., Tagscherer, U., & Chang, A. Y. Beyond catch-up—can a new innovation policy help China overcome the middle income trap?. *Science and Public Policy*. 2017; *44*(5): 656-669. doi:https://doi.org/10.1093/scipol/scw092

8. http://data.uis.unesco.org/ (accessed on 31 July 2021).

9. https://www.cnipa.gov.cn/col/col61/index.html#mark (accessed on 31 July 2021).

10. Cornell University, INSEAD, and WIPO. *The Global Innovation Index: Who will Finance Innovation?* Ithaca, Fontainebleau, and Geneva. 2020. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2020.pdf

11. State Council of the People's Republic of China and The Central Committee of Communist Party of China (CPC), 2016. National Strategy of Innovation-Driven Development http://www.gov.cn/gongbao/content/2016/content_5076961.htm

12. Seiford, L. M., & Zhu, J. Profitability and marketability of the top 55 US commercial banks. *Management science*. 2019; 45(9): 1270-1288. doi: https://doi.org/10.1287/mnsc.45.9.1270

13. Kao, C., & Hwang, S. N. Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. European journal of operational research. 2008; 185(1): 418-429. doi:https://doi.org/10.1016/j.ejor.2006.11.041

14. Kao, C. Network data envelopment analysis: A review. *European journal of operational research*. 2014; *239*(1): 1-16. doi:https://doi.org/10.1016/j.ejor.2014.02.039

15. Färe, R., Grabowski, R., Grosskopf, S., & Kraft, S. Efficiency of a fixed but allocatable input: A non-parametric approach. *Economics Letters*. 1997; 56(2): 187-193. doi: https://doi.org/10.1016/S0165-1765(97)81899-X

16. Yang, F., Jiang, L., & Ang, S. A winner-take-all evaluation in data envelopment analysis. *Annals of Operations Research*. 2019; 278(1): 141-158.

doi:https://doi.org/10.1007/s10479-018-2833-z

17. Ang, S., Liu, P., & Yang, F. Intra-Organizational and inter-organizational resource allocation in two-stage network systems. *Omega*. 2020; 91: 102009. doi:https://doi.org/10.1016/j.omega.2018.11.018

18. Min, S., Kim, J., & Sawng, Y. W. The effect of innovation network size and public R&D investment on regional innovation efficiency. *Technological Forecasting and Social Change*. 2020; *155*: 119998. doi:https://doi.org/10.1016/j.techfore.2020.119998

19. Chen, K., & Guan, J. Measuring the efficiency of China's regional innovation systems: application of network data envelopment analysis (DEA). *Regional Studies*. 2012; *46*(3): 355-377. doi:https://doi.org/10.1080/00343404.2010.497479

20. Xiong, X., Yang, G. L., & Guan, Z. C. A parallel DEA-based method for evaluating parallel independent subunits with heterogeneous outputs. *Journal of Informetrics*. 2020; *14*(3): 101049. doi:https://doi.org/10.1016/j.joi.2020.101049

21. Kao, C. Efficiency decomposition in network data envelopment analysis: A relational model. *European journal of operational research*. 2009; *192*(3): 949-962. doi:https://doi.org/10.1016/j.ejor.2007.10.008

22. Jianfeng, Ma. A two-stage DEA model considering shared inputs and free intermediate measures. *Expert Systems with Applications*. 2015; *42*(9): 4339-4347. doi: https://doi.org/10.1016/j.eswa.2015.01.040

23. Simeth, M., & Raffo, J. D. What makes companies pursue an open science strategy?. *Research Policy*. 2013; *42*(9): 1531-1543.

24. Soule, E. K., Chaiton, M., Zhang, B., Hiler, M. M., Schwartz, R., Cohen, J. E., & Eissenberg, T. Menthol cigarette smoker reactions to an implemented menthol cigarette ban. *Tobacco Regulatory Science*, 2019:*5*(1), 50-64. doi: https://doi.org/10.18001/TRS.5.1.5

25. Pelletier, H., Diemert, L., Soule, E., Cohen, J. E., Eissenberg, T., O'Connor, S., & Schwartz, R. Smokers' Experiences with Vaping to Quit: A Concept Mapping Study. *Tobacco Regulatory Science*,2021; *7*(2), 103-117. doi: https://doi.org/10.18001/TRS.7.2.2