Analysis of Input and Output of China's Agriculture Based on Canonical Correlation

CHEN Dao-ping *
School of Economics and Management, Chongqing Normal University, Chongqing 400047, China

Abstract I select effective irrigated area, consumption of agricultural chemical fertilizer, electricity consumed in rural areas, and total power of agricultural machinery as input variables of China's agriculture; I select grain, bean, tobacco, oil-bearing crop and fruit as output variables of China's agriculture. By using the data of *China Statistical Yearbook in* 2010, based on the analysis method of canonical correlation, I conduct research on the input and output of China's agriculture. The results show that consumption of chemical fertilizer has the biggest impact on the agricultural output of China, followed by the input of total power of agricultural machinery; the canonical variable of agricultural output of China is mainly impacted by grain, oil-bearing crop and fruit; in terms of the selected variables, the output increase of grain, oil-bearing crop and fruit in China arises from the input increase of agricultural chemical fertilizer and machinery, and there is high-degree correlation between the two. According to the conclusions, the policy suggestions are put forward as follows: gradually decrease consumption of chemical fertilizer; increase the use of modern agricultural machinery; increase agricultural irrigation input.
Key words Agriculture, Input, Output, Canonical correlation, China

Since the reform and opening up, the agriculture of China has made a great progress. Be that as it may, the agriculture of China is also confronted by the problems of deterioration of agricultural ecological environment and quality safety of agricultural products. One of important factors responsible for these problems is the problem of agricultural input structure. Putting the limited capital in the construction of water conservancy, purchase of equipments, the use of mechanical force, or the increase of chemical fertilizer use is a difficult question for us to make decision: increasing consumption of chemical fertilizer can get instant effects of augmenting the agricultural yield, but it may cause ecological problems and quality problems of agricultural products; while constructing agricultural irrigation facilities or developing innovative modern agricultural machinery plays unsatisfactory role in promoting the agricultural yield. In order to solve the problem, we should at first probe into the current input situation of these influencing factors of agricultural output, the impact of these factors on output and the mutual relationship between input and output. In local areas, on the basis of simple correlation analysis and linear regression, Shi Mingying conducts analysis on the agricultural input and output, and thinks that the total power of agricultural machinery has the greatest impact on agricultural total output value, and chemical fertilizer is the main factor impacting the output growth of grain[1]; Kong Chaoli uses principal component analysis method to research the agricultural input of China, and thinks that water-saving and fertilizer-saving precise agriculture will become the important direction of agricultural development of China in the future[2]. Based on the canonical correlation analysis method, I conduct research on the mutual relationship between input and output of China's agriculture, so as to provide theoretical basis for decision making of agricultural input structure in China.

Received: June 13, 2011 Accepted: July 2, 2011
Supported by Scientific Research Program of Chongqing Municipal Commission of Education (KJ110619); Ph. D. Program Foundation of Chongqing Normal University (11XWB004).
* Corresponding author. E-mail: cdp8cdp@163.com

1 Index selection, data source and research method

1.1 Index selection

1.1.1 Index of agricultural input. The agricultural input is measured by four indices.

1.1.1.1 Effective irrigated area (*GGMJ*, $\times 10^3$ hm^2). Effective irrigated area refers to the area of farmland that can be irrigated regularly in the year, with certain water source, level ground, and sound supporting irrigation projects or equipments. Under normal circumstances, the effective irrigated area should be equal to the area summation of paddy field and irrigable land that can conduct normal irrigation when they are outfitted with irrigation projects or equipments, which is an important indicator of reflecting drought resistance capacity of China's farmland. The effective irrigated area also reflects the amount of irrigation inputs.

1.1.1.2 Use of agricultural fertilizer (*HFYL*, 10 000 t). Consumption of chemical fertilizers is the actual amount of chemical fertilizers that are used for agricultural production in one year, including nitrogen, phosphorus, potash and compound fertilizer. The using amount of chemical fertilizer is required to be calculated according to standard consumption. Standard consumption is the amount of nitrogen, phosphate, and potash after being converted, according to the percentage of their respective main component, namely nitrogen, phosphorus pentoxide, and potassium oxide; compound fertilizer is converted according to its main component, and the formula is as follows: Standard consumption = Consumption of chemical fertilizer × Percentage of the amount of certain effective ingredient of fertilizer.

1.1.1.3 Electricity consumed in rural areas (*NYDL*, 0.1 billion kW · h). Electricity consumed in rural areas includes electricity for power and electricity for living in rural areas, excluding the electricity for county industries and electricity for urban life.

1.1.1.4 Total power of agricultural machinery (*JXDL*, 10 000 kW). The total power of agricultural machinery includes farm machinery, irrigation machinery, harvesting machinery, and agricultural transportation machinery (internal combustion engine is calculated after the horsepower is converted into watt, and electric motor is calculated after the power is converted into watt), excluding the power machinery and operating machinery that are specifically used for industries set up by township and village, infrastructure, non-agricultural transport, scientific experiment, teaching and other non-agricultural production aspects.

1.1.2 Index of agricultural output. Agricultural output is measured by five indices.

1.1.2.1 Output of grain (*GL*, 10 000 t). Grains mainly includes rice, wheat, corn and so on, which are produced by the farms set up by state-owned economic bodies, collective economic bodies, rural households, and industrial and mining enterprises as well as other production departments.

1.1.2.2 Output of bean (*DL*, 10 000 t). Beans are calculated according to the dry beans without pods.

1.1.2.3 Output of tobacco (*YY*, 10 000 t). Tobacco mainly includes flue-cured tobacco.

1.1.2.4 Output of oil (*YL*, 10 000 t). Output of oil refers to the output of all oil crops, including peanut, rapeseed, sesame, sunflower seed, flaxseed and other oil-bearing crop, but does not include soybeans, woody oleiferous plants and wild oil-bearing crop. Peanut is calculated according to the dry peanut with hull.

1.1.2.5 Output of fruit (*SG*, 10 000 t). Fruits include apples, oranges, pears, grapes, bananas, and so on.

1.2 Data source The study uses the data concerning the aforesaid indices from 31 provinces, municipalities directly under the jurisdiction of the Central Government and autonomous regions in 2009, and some of the data are from the agricultural part of *China Statistical Yearbook* in the year 2010.

1.3 Research method We use the principle of canonical correlation to analyze the input and output of agriculture in China. Let the random vector $x = (x_1, x_2, \cdots, x_p)$, $y = (y_1, y_2, \cdots, y_q)$, then the correlation between x and y is called Canonical Correlation. When $p = 1$ and $q = 1$, it is simple correlation; when $p = 1$ and $q \neq 1$ or $p \neq 1$ and $q = 1$, it is multiple correlation. The simple correlation and multiple correlation are the special cases of canonical correlation. Canonical correlation analysis was advanced by Hotelling in the year 1936[3-4]. The principle of canonical correlation analysis is that we extract one canonical variable u_1 from the first group of x, and extract one canonical variable v_1 from the second group of y, namely that:

$$\begin{cases} u_1 = a_{11}x_1 + a_{21}x_2 + \cdots + a_{p1}x_p \\ v_1 = b_{11}x_1 + b_{21}x_2 + \cdots + b_{q1}x_q \end{cases}$$

We make the correlation coefficient $\rho(u_1, v_1)$ of u_1 and v_1 maximum, then extract canonical variable u_2 from x and extract canonical variable v_2 from y, namely that:

$$\begin{cases} u_2 = a_{12}x_1 + a_{22}x_2 + \cdots + a_{p2}x_p \\ v_2 = b_{12}x_1 + b_{22}x_2 + \cdots + b_{q2}x_q \end{cases}$$

We make the correlation coefficient $\rho(u_2, v_2)$ of u_2 and v_2 secondly biggest, namely $\rho_1^2(u_1, v_1) \geqslant \rho_2^2(u_2, v_2)$, and u_2, v_2, and u_1, v_1, are mutually independent, but u_2 correlates with v_2; in the similar manner, finally we get k couples of canonical variable u_k, v_k, $k \leqslant \min(p, q)$ until the correlation between two variables are extracted all. When one sample of random vector x and y is known, coefficient a_{ij} ($i = 1, 2, \cdots, p$; $j = 1, 2, \cdots, k$) and b_{ij} ($i = 1, 2, \cdots, q$; $j = 1, 2, \cdots, k$) can be calculated[2], so as to get canonical correlation model. We can further conduct analysis on the canonical structure, that is, analyze the mutual relationship between canonical variable and original variable.

2 Result and analysis

2.1 Establishment of canonical correlation model According to the principle of canonical correlation analysis, by using Canonical Correlation Analysis program of SPSS13.0[5], we get the canonical correlation coefficient and significance test result, which can be seen in Table 1.

Table 1 Canonical correlation coefficient and the significance test

Canonical variable	Correlation coefficient	Wilk's	Chi-SQ	Degrees of freedom	Probability
(u_1, v_1)	0.984	0.015	84.265	20	0.000
(u_2, v_2)	0.587	0.464	15.361	12	0.222
(u_3, v_3)	0.534	0.708	6.913	6	0.329
(u_4, v_4)	0.103	0.989	0.212	2	0.899

Table 1 shows that the correlation coefficient of the first pair of canonical variables is 0.984, and the probability is much less than the significance level of 0.01, so the correlation coefficient of the first pair of canonical variables (u_1, v_1) is statistically significant at the significance level of 0.01; the correlation coefficient of the second pair of canonical variables (u_2, v_2) is not statistically significant at the significance level of 0.05; the correlation coefficient of the third pair of canonical variables (u_3, v_3) is not statistically significant at the significance level of 0.05; the correlation coefficient of the fourth pair of canonical variables (u_4, v_4) is not statistically significant at the significance level of 0.05. So we choose the first pair of canonical variables to conduct analysis afterwards.

The program of Canonical Correlation Analysis offers the standardized coefficient of canonical variable, which can be shown in Table 2, 3 respectively.

Table 2 Canonical variable coefficient of input

Input	u_1	u_2	u_3	u_4
GGMJ	−0.053	0.363	−2.128	−0.154
HFYL	−0.745	−1.657	0.744	−0.819
NYDL	0.068	0.686	0.297	−0.759
JXDL	−0.258	1.240	1.170	1.246

Table 3 Canonical variable coefficient of output

Output	v_1	v_2	v_3	v_4
GL	−0.526	1.303	−0.201	0.118
DL	0.035	−0.588	−0.819	0.301
YY	−0.042	−0.327	0.364	0.867
YL	−0.248	−1.681	−0.012	−0.896
SG	−0.349	0.523	0.317	0.661

According to Table 2, 3, we can get the canonical correlation model of statistically significant canonical variables as follows:

$$\begin{cases} u_1 = 0.053GGMJ + 0.745HFYL - 0.068NYDL + 0.258JXDL \\ v_1 = 0.526GL - 0.035DL + 0.042YY + 0.248YL + 0.349SG \end{cases}$$

From the coefficient of canonical variable of the model, we can find that consumption of chemical fertilizer (*HFYL*) has the greatest impact on the canonical variable u_1, followed by the total power of agricultural machinery (*JXDL*), but the total power of agricultural machinery has far too small impact on u_1 in comparison with consumption of chemical fertilizer; effective irrigated area (*GGMJ*) and electricity consumed in rural areas (*NYDL*) have small impact on the canonical variable u_1, so we think that u_1 mainly refects the the input of chemical fertilizer in agriculture. In terms of output, grain (*GL*), fruit (*SG*) and oil-bearing crop (*YL*) have relatively big impact on the canonical variable u_2 that reflects output, while bean (*DL*) and tobacco (*YY*) are not the main influencing factors.

2.2 Analysis of canonical structure Canonical structural analysis is to use the correlation coefficient between canonical variables and original variables to analyze the correlation among the variables at great length, which can help us to understand the structure among variables.

Table 4 Correlation coefficient of canonical variable and original variable

Canonical variable	*GGMJ*	*HFYL*	*NYDL*	*JXDL*	*GL*	*DL*	*YY*	*YL*	*SG*
u_1	0.884	0.986	0.258	0.915	0.908	0.135	0.122	0.888	0.820
v_1	0.870	0.970	0.254	0.900	0.922	0.138	0.124	0.902	0.834

It can be seen from Table 4 that on the one hand, the canonical variable u_1 has high correlation with the input variables of consumption of chemical fertilizer (*HFYL*), the total power of agricultural machinery (*JXDL*), and effective irrigated area (*GGMJ*), while the correlation coefficient of consumption of chemical fertilizer (*HFYL*) and the canonical variable u_1 is the greatest of 0.986; in the mean time, consumption of chemical fertilizer (*HFYL*) and the total power of agricultural machinery (*JXDL*) have great correlation with the canonical variable v_1, and the correlation coefficient is greater than 0.900. On the other hand, the canonical variable u_1 has strong correlation with the output variables of grain (*GL*), fruit (*SG*) and oil-bearing crop (*YL*). The canonical variable v_1 also has strong correlation with the output variables of grain (*GL*), fruit (*SG*) and oil-bearing crop (*YL*). From the two aspects of correlation, we can know that the agricultural input in China is reflected in agricultural chemical fertilizer and power of agricultural machinery, while the agricultural output in China is reflected in grain, fruits and oil-bearing crop, that is, the increase of output of grain, fruits and oil-bearing crop in China arises from the increase of input of chemical fertilizer and power of agricultural machinery.

3 Conclusion and policy suggestions

3.1 Conclusion This paper selects effective irrigated area, consumption of chemical fertilizer, electricity consumed in rural areas, and total power of agricultural machinery as input variables of China's agriculture; this paper selects grain, bean, tobacco, oil-bearing crop and fruit as output variables of China's agriculture. By using the data of *China Statistical Yearbook* in 2010, based on the analysis method of canonical correlation, this paper conducts research on the input and output of China's agriculture. The results show that consumption of chemical fertilizer has the biggest impact on the agricultural output of China, followed by the input of total power of agricultural machinery; the canonical variable of agricultural output of China is mainly impacted by grain, oil-bearing crop and fruit; in terms of the selected variables, the output increase of grain, oil-bearing crop and fruit in China arises from the input increase of agricultural chemical fertilizer and machinery, and there is high-degree correlation between the two.

3.2 Policy suggestions According to the above conclusion, the corresponding policy suggestions are put forward as follows.

3.2.1 Gradually reduce consumption of chemical fertilizer. Amid several input factors, chemical fertilizer is the most important input factor. Moreover, chemical fertilizer has a great impact on the necessaries of life of China's residents, such as grain, oil-bearing crop and fruit. This may have negative impact on health of China's residents, and considerable consumption of chemical fertilizer, in the long run, will also have negative impact on soil and ecology.

3.2.2 Increase the use of mechanical force and develop modern agricultural machinery. As the power of machinery has important influence on the output of agriculture, so after decreasing consumption of chemical fertilizer gradually, one way to maintain the increase of agricultural output is to use more power of machinery in agricultural production.

3.2.3 Construct irrigation project and increase input in irrigation equipments. From the preceding analysis, we can know that the importance of irrigation in agricultural inputs has not yet well been reflected. After decreasing the consumption of chemical fertilizer gradually, in order to continue to maintain the increase of agricultural output, we can adopt the way of increasing irrigation input.

References

[1] SHI MY, ZHU YJ. Analysis of effects of agricultural input-output in Anhui Province[J]. Operations Research and Management, 1999 (2): 62–66. (in Chinese).

(To page 15)

Continued (Table 6)

Region	Score of F_1	Score of F_2	Score of F_3	Score of F_4	Overall Score	Overall ranking
Henan	−0.760 3	0.828 0	1.847 6	0.014 6	0.045 3	13
Hubei	−0.322 9	−0.186 8	0.628 4	0.836 8	−0.084 1	15
Hunan	−0.764 9	0.695 5	1.453 5	0.486 7	−0.019 4	14
Guangdong	0.864 6	−0.474 1	0.997 5	0.821 3	0.562 6	6
Guangxi	−1.399 5	0.226 9	−0.260 5	1.140 9	−0.698 6	28
Hainan	−2.425 4	1.364 9	−1.906 8	1.918 5	−1.196 0	31
Chongqing	0.266 0	−1.439 3	0.338 2	0.493 9	−0.116 7	16
Sichuan	−0.564 3	−0.894 5	1.390 3	0.270 2	−0.308 2	22
Guizhou	−0.536 7	−1.444 9	−0.571 2	0.380 8	−0.705 4	29
Yunnan	−1.032 6	−0.951 1	−0.375 2	0.077 3	−0.852 8	30
Tibet	−0.682 8	0.669 9	−0.898 3	−3.441 3	−0.550 9	26
Shaanxi	−0.163 9	−0.729 0	−0.581 9	0.451 8	−0.324 4	24
Gansu	−0.477 3	−0.887 4	−1.154 6	0.325 6	−0.628 2	27
Qinghai	0.081 1	−0.745 5	−0.985 0	−0.793 4	−0.322 6	23
Ningxia	−0.019 3	−0.221 9	−0.779 9	−1.150 3	−0.244 3	20
Xinjiang	−0.864 5	0.686 2	−0.919 1	−0.653 2	−0.490 8	25

2.3 Cluster analysis On the basis of factor analysis, we conduct sample clustering. According to the scores of factors, we divide 31 regions into 4 types which can better reflect the actual situation of China. The classification of 31 regions can be seen in Table 7. The cluster results directly reflect the similarity and difference of level of the agricultural economic development in 31 cities of China.

Table 7 Cluster result of regions

Layer of rural economic level	Overall mark	Evaluation of rural economic level	Region
The first layer	More than 0.9	The regions with developed agricultural economy	Beijing, Shanghai, Zhejiang, Tianjin, Jiangsu
The second layer	0 −0.9	The regions with relatively developed agricultural economy	Guangdong, Shandong, Fujian, Liaoning, Hebei, Shanxi, Jiangxi, Henan
The third layer	−0.4 −0	The regions with less-developed agricultural economy	Hunan, Hubei, Chongqing, Anhui, Heilongjiang, Jilin, Ningxia, Inner Mongolia, Sichuan, Qinghai, Shaanxi
The fourth layer	Below −0.4	The regions with underdeveloped agricultural economy	Xinjiang, Tibet, Gansu, Guangxi, Guizhou, Yunnan, Hainan

3 Conclusion and suggestions

Through multivariate statistical analysis, the 4 types of economic development we have classified, to much extent, are consistent with the actual situation of development of all provinces and regions of China in reality we have known. We can clearly find that the level of agricultural economic development in all regions of China is uneven; the level of rural economic development in all provinces are consistent with the overall level of economic development, which verifies that the rural economy plays the role of supporting the national economy; the level of rural economic development has obvious geographical differences, and the rural economic development of the central and eastern China has obvious advantages in comparison with the rural economic development of the western China.

There are multifarious factors responsible for the differences of agricultural economy in provinces and regions of China. China should pay attention to these regional characteristics, formulate the relevant policies, focus on solving the imbalance of rural economic development, adjust the agricultural economic level by steps with purposefulness, spare no efforts to bridge the regional gap, and support the rural economic development of the northwestern regions of China, so that we achieve the overall goal of joint development and common prosperity.

References

[1] Rural Social Economic Survey Investigations Division, National Bureau of Statistics of China. China Rural Statistical Yearbook[M]. Beijing: China Statistics Press, 2008 (in Chinese).
[2] WU J, LI YX. Analysis about the developing level of agro−moderuization in various prefectures of Xinjiang[J]. Xinjiang State Farms Economy, 2009, 7: 37−41 (in Chinese).
[3] YU XL, REN XS. Multivariate Statistical Analysis[M]. Beijing: China Statistics Press, 1999. (in Chinese).
[4] JIANG K, YANG MY, ZHOU S, *et al*. Multivariate statistics analysis in cities of Fujian Province[J]. Taiwan Agricultural Research, 2008, 3: 58−61 (in Chinese).
[5] CHEN GJ, YUAN L, LIU Y. Analysis of agricultural competitiveness factor in Jiangsu area[J]. Contemporary Economics, 2007, 9: 94−95. (in Chinese).

(From page 11)
[2] KONG CL, YANG QC. The regression models of principle component analysis on in-output variables in agriculture[J]. Journal of Anshan Normal University, 2006(8): 11−14. (in Chinese).
[3] HE XQ. Multivariate Statistical Analysis[M]. Beijing: China Renmin University Press, 2004. (in Chinese).
[4] ZHANG RT, FANG KT. Introduction Multivariate Statistical Analysis[M]. Beijing: Science Press, 2006. (in Chinese).
[5] ZHANG WT. Statistical Analysis Tutorial-advanced articles[M]. Beijing: Beijing Hope Electronic Press, 2002. (in Chinese).