A Comprehensive Analysis of China Regional Agriculture Based on Efficiency, Advantage and Development Strategy*

Yong Liu[1] and Hong Zhang[2]

[1]College of Literature Law & Economics
[2]College of Computer Science & Technology
[1, 2] Wuhan University of Science & Technology, Wuhan, China

Abstract

Agriculture is a traditional weak industry in most minds, so fiscal support is necessary for most countries and regions, but result in different efficiency. For countries like China with such a huge area, agriculture development for different region is extremely imbalanceable with different comparative advantage. This paper first gauges the efficiency value of agriculture fiscal support with the well-known Panel-Data Model for the dataset of 25 provinces in China in recent 25 years. Then we calculate agriculture advantage values of these 25 provinces from two angles—absolute advantage and comparative advantage. During discussion part, we comprehensively consider efficiency and comparative advantage of each province. And we try to classify these 25 provinces into 3 clusters after comprehensive considerations of the above aspects. Consequently, according to clustered results, the proposal for China agriculture development proposed by this paper will be that different province should be with different agriculture development strategy.

Keywords: *Efficiency, Advantage, Strategy, Panel-data model, Cluster Analysis*

1. Introduction

For agriculture is a traditional weak industry in most minds, so almost all countries and regions pay special attentions on its development, and give large amount of finance support. So, many researchers focused on the relations between finance support and agriculture economy. Li Huanzhang, Qian Zhonghao [1] analyzed the relation between China finance support and agricultural output and results in mutual Granger Causality for them based on Granger Causality Test; Shen Xiaoli, Anlongsong [2] analyzed relations between finance support and agricultural economy increase with modified C-D Production Function for Hu'nan Province in China, resulted in finance support supplying 46.73% contribution of agricultural output; Wei Lang [3,4] analyzed finance support of China and its western part, and found that there are obvious different efficiency between regions.

While some other researchers study agriculture based on comparative advantage, Lan Wanlian [5] designed series of index and ranked comparative advantage for all provinces in China; Yu Zhanping [6] considered agricultural comparative advantage of industry structure, industry organize and industry benefit, assessed, and experience-analyzed agriculture comparative advantage of Tianjin in China; Huang Zhaoying [7] experience-analyzed whole agriculture comparative advantage of China with indicative comparative advantage index and resulted in China's agriculture advantage gradually- weakened.

Thirdly, Huang Dajin (2007), Wang Enhu (2007), Yu Yanchun (2007) focused on the agriculture development strategy from different angels, such as food consumption, energy and so on.

During economy development, comparative advantage is usually the primary consideration. Also for a country's development, comparative advantage is very important, which has been proved by the successful experiences of East Asian countries and some less-developed countries [9, 10, 11].

Fiscal support efficiency, comparative advantage and development strategy are all hot research topics in the area of agriculture economy. But most researches just only observe one point and ignore other parts, few researches have focused on comprehensive information of efficiency, comparative advantage and development strategy. In fact, to analyze such comprehensive information with mathematical models and statistical techniques could be more credible and valuable for agriculture development. The main challenge is how to integrate these important factors: efficiency and advantage and strategy. For example, different provinces in China have different agriculture environment, in other words, different comparative advantage, while their fiscal-supports to agriculture result in different efficiencies.

The rest of this paper is organized as follows: in Section II, Panel-Data Model is introduced; in Section III, Panel-Data Model is used to estimate the efficiencies of agriculture fiscal-support for 25 provinces in China; in Section IV, Advantage values of all provinces in China are calculated; in Section V k-means clustering method is introduced,; in Section VI, k-means clustering method is employed on the data of efficiency, comparative advantage, first respectively, then together; and lastly in section VII conclusion is deduced.

2. Data Sources and Modeling Assumption

In this paper Panel-Data Model is used to gauge the efficiency of fiscal support to agriculture of 25 provinces in China, dataset are from "CHINA COMPENDIUM OF STATISTICS 1949~2004. Other six provinces as Anhui, Hainan, Chongqing, Sichuan, Gansu, Qinghai and regions as Taiwan, Macao and Hongkong are not studied in this paper for their related dataset are incomplete or could not be obtained.

The reason why Panel-Data Model is adopted is that there are so many objectives—25 provinces needed study, and each objective has 25 years data from 1980 to 2004, so Panel-Data Model is the best choice for analysis.

2.1 Panel Data Model

Panel-Data Model is also called TS/MS(Time series/ Cross series) model, the model needs investigating period of same kind of cross-section data, including three-dimensional data structure as objectives, times, and index. It could also be understood as getting many cross-sections in time series. Sample dataset from these cross-sections are chosen simultaneously for model building, comparing with purely time series data or purely section data. This kind of panel dataset could provide more sample data, improve estimation validity, and analyze problem more completely.

Suppose dependent variable Y and $1\times k$ dimensional observed explanatory variable X_{it}, meet linearity relationship in equation (1):

$$Y_{it} = \alpha_i + \beta_i X_{it} + \mu_{it} \tag{1}$$

in which, i=1, 2, L, N; t=1, 2, L, t. the model will consider relationships of k economic parameters of N units at T time points, N is number of cross-sections, T represents observed times; α_{it} denotes constant parameter, $\beta_{it} = (\beta_{1t}, \beta_{2t}, \ldots \beta_{kt},)$ represents related coefficient of $k \times 1$ dimension explanatory variable $Xit = (x_{1t}, x_{2t}, \ldots, x_{kt})$, k denote number of explanatory variables, random errors μ_{it} are dependent mutually, and are supposed to be zero mean value, same variance. Though the form of Panel-Data Model is shown as equation (1), it could be divided into 4 kinds according to parameter conditions:

- Model I is constant value model, that is: $\alpha_i = \alpha_j, \beta_i = \beta_j$, this kind of model has no special influence on cross-section unit, and no structure change, like putting cross-section data at several periods together as sample data;
- Model II is changeable intercept model, that is: $\alpha_i \neq \alpha_j, \beta_i = \beta_j$, this kind of model has special influence on cross-section units;
- Model III is changeable slope model, that is: $\alpha_i = \alpha_j, \beta_i \neq \beta_j$, this kind of model has special influence on cross-section units in periods;
- Model IV is changeable coefficient model, that is : $\alpha_i \neq \alpha_j, \beta_i \neq \beta_j$, this model has not only special influence on cross-section units, but also influence in periods.

Model II, III and IV have special influences on cross-section units, these influences, which are divided into fixed influences and random influences, are caused from ignored variables of each unit reflecting their economic differences.

2.2 Verifying of Specific Model

With panel data for model building, it is primary to verify the specific form of the model that is to verify whether the model parameters are same in cross-sections and in periods or not, verifying process follows the hypothesis listed below:

Hypothesis 1:

intercepts and slopes are same in cross-section and periods, that is:

null hypothesis , $H10$: $\alpha_i = \alpha_j, \beta_i = \beta_j$

alternative hypothesis: $H11$: $\alpha_i \neq \alpha_j or \beta_i \neq \beta_j$

Hypothesis 2:

slopes are same in cross-section and in periods, while intercepts are different, that is:

null hypothesis $H20$: $\alpha_i \neq \alpha_j, \beta_i = \beta_j$

alternative hypothesis $H21$: $\alpha_i \neq \alpha_j, \beta_i \neq \beta_j$, or $\alpha_i = \alpha_j, \beta_i \neq \beta_j$

Hypothesis 3:

intercepts are same in cross-section and periods, while slopes are different, that is:

null hypothesis $H30$: $\alpha_i = \alpha_j, \beta_i \neq \beta_j$

alternative hypothesis $H31$: $\alpha_i \neq \alpha_j$且$\beta_i \neq \beta_j$

Obviously, if hypothesis 1 is accepted, no more verifying is needed, model I could be adopted directly; if hypothesis $H10$ is rejected, then hypothesis 2 should be verified, if $H20$ is accepted, model II will be adopted; if $H20$ is rejected, hypothesis 3 will be verified. If $H30$ is accepted, model III would be adopted, or model IV will be the only choice.

Verifying method will be covariance verifying, and specific model type will be decided according to the comparison of statistical F value computed as following equation3, 4, or 5 with related critical F value, step by step:

$$F_1 = \frac{(S_1 - S_4)/(N-1)(k+1)}{S_4/[N(T-k-1)]} \sim F[(N-1)(k+1), N(T-k-1)] \quad (2)$$

$$F_2 = \frac{(S_2 - S_4)/(N-1)\,k}{S_4/[N(T-k-1)]} \sim F[(N-1)\,k, N(T-k-1)] \quad (3)$$

$$F_3 = \frac{(S_3 - S_4)/(N-1)}{S_4/[N(T-k-1)]} \sim F[(N-1), N(T-k-1)] \quad (4)$$

in which, S_1, S_2, S_3, S_4 are respectively denote residuals of model I, II,III and IV during estimation,; N is number of cross-sections, T is number of periods; k is number of parameters for estimation (not including intercepts). If value F_1 computed as left in equation 2 is less than critical F value in the right of equation 2 from standard sheet, model I should be adopted; or to the next step, same operations to equation 3 and 4 one by one. And lastly model is the only choice if F_1, F_2 and F_3 are all not qualified to judge rules.

To the study in this paper, firstly, specific Panel-Data Model type would be selected according to above F test. Panel-Data Model will be decided by these computation values: S_1=2007.337 , S_2=73.6734 , S_3=1720.697 , S_4=57.9419 , so F_1=403.0607 , F_2=6.5057 , F_3=711.550 , after checking F value distribution sheet, under 5% significance, critical values are: $F_{1,0.05} = F(48,575) = 1.39$, $F_{2,0.05} = F(24,575) = 1.52$, $F_{3,0.05} = F(24,575) = 1.52$, so under this significance, hypothesis 1 to hypothesis 3 are all rejected, model in this paper should be changeable intercept and changeable coefficient, that is model , it means for each province and each year relationship between fiscal supporting and agriculture growth are obvious different.

Secondly, under this foundation, fixed effects model or random effects model will be selected according to Hausman test, calculated value of statistical value W=3996.43, so the model in this paper should be fixed effects type model.

3. Agriculture Fiscal Support Efficiency Analyzing Results

In this section, selected Panel-Data Model through section II is used to estimate the efficiency of fiscal support to agriculture of each province in China. In the equation 5 Value represents the total value (one hundred million Yuan) of agriculture, forestry, animal, husbandry and fishing ("AFAHF" for short). Fiscal represents fiscal support value (one hundred million Yuan) for AFAHF. To eliminate Heteroscedasticity during estimation, Logarithmic transformation to these two variables is necessary, such measures will not change their inherent relationship, new variables are log(value) and log(fiscal). Detail model form as following equation 5:

$$\log(value_i) = \alpha^* + \alpha_i + \beta_i \log(fiscal_i) \qquad (5)$$

Fitting results with Fixed effect changeable coefficient model are shown in Table 1.

There are 6 columns in the table, column 1 is the name of the 25 provinces of China studied in this paper; α_i column is the fixed effects, while value α^* is the common effects for all provinces; other values below α_i except α^* reflecting special effect of each province (such as Beijing -0.289) for their unique characteristics ; column 3 is the estimation value of coefficient β_i in the model; Column 4 is σ represents the standard error of β_i, column 5 is T-value and column 6 is P-value are all for the judgment of each equation; i denotes different province.

Furthermore, in the table, α_i column represents influence on AFAHF from economic factors other than fiscal supports; while β_i column reflects the only fiscal supporting efficiency to AFAHF.

For dependent variable in this paper (value) and explanatory variable (fiscal) are Logarithmic-transformed, the coefficient β_i is actually the elasticity between local agriculture fiscal support and local AFAHF growth, so, we can conclude that, if there is one percent growth of agriculture fiscal support value, the AFAHF value will be growing at β_i percent rate. We can also call β_i as efficiency of e agriculture fiscal support.

From table 1, it is obvious that the model fitting degree is excellent, not only for each province T-value (column 5) and P-value (column 6) pass the test, but also as a whole in table 1 the bottom 6 rows, which reflects the comprehensive assessment for the equations, such as Prob (F-statistic) is 0.000 and R-squared is 0.948 (adjusted R-squared 0.943), fitting degree is significant. it is also shown that there is high correlation between fiscal support and AFAHF growth.

Table 1. Estimation results

Variable log(fiscal)	α_i	β_i	σ	**T-value**	**P-value**
α^*	3.446		0.032	108.212	0.00
Beijing	-0.289	0.800	0.054	14.864	0.00
Tianjin	-0.284	1.20	0.077	15.542	0.00

Hebei	1.181	0.869	0.056	15.519	0.00
Shanxi	-0.040	0.859	0.071	12.026	0.00
Neimeng	-1.158	1.308	0.088	14.833	0.00
Liaoning	-0.179	1.11	0.074	14.983	0.00
Ji'lin	-0.147	1.134	0.078	14.476	0.00
Heilongjiang	-0.235	0.972	0.077	12.605	0.00
Shanghai	-0.379	0.961	0.068	14.037	0.00
Jiangsu	0.856	0.889	0.064	13.967	0.00
Zhejiang	0.744	0.820	0.062	13.123	0.00
Fujian	-0.167	1.22	0.070	17.366	0.00
Jiangxi	1.260	0.792	0.058	13.710	0.00
Shandong	0.845	0.923	0.062	14.915	0.00
Henan	0.032	1.233	0.079	15.656	0.00
Hubei	1.128	0.800	0.080	9.947	0.00
Hu'nan	0.367	0.951	0.069	13.787	0.00
Guangdong	1.018	0.745	0.050	14.879	0.00
Guangxi	-0.163	1.205	0.080	15.139	0.00
Guizhou	-0.176	1.172	0.101	11.560	0.00
Yun'nan	-0.457	0.930	0.061	15.150	0.00
Xizang	-0.772	0.811	0.076	10.661	0.00
Shaan'xi	0.010	0.854	0.067	12.794	0.00
Ningxia	-1.889	0.930	0.065	14.257	0.00
Xijiang	-1.106	1.423	0.092	15.456	0.00
Effects Specification					
Cross-section fixed (dummy variables)					
R-squared		0.948	Mean dependent var		5.417

Adjusted R-squared	0.943	S.D. dependent var	1.335
S.E. of regression	0.317	Akaike info criterion	0.620
Sum squared resid	57.942	Schwarz criterion	0.975
Log likelihood	-143.614	F-statistic	213.460
Durbin-Watson stat	0.514	Prob (F-statistic)	0.000

Note: the model is fitted with Eviews 5.0 software

4. Advantage of Agriculture

In section III, agriculture fiscal support efficiency is gauged through Panel-data model. All provinces support agriculture with fiscal capital for its great importance and also show their special focuses on agriculture development. But during economy development, comparative advantage is often the primary consideration, so we should not ignore agriculture advantage analysis for each province.

4.1. Comparative Advantage one: APCA-- Agriculture production comparative advantage

Comparative advantage theory originated from David Ricardo only from single labor-cost analysis and it is too simple for complex economy activity; later researchers supplement and revise comparative advantage theory from multi-aspects such as Gottfried Von Haberler proposed opportunity cost to substitute labor-cost analysis to explain comparative advantage, and Heckscher－Ohiln from factor endowment angle to advocate comparative advantage thought. We rank APCA by increasing order in table 3.

Table 2. APCA--Agriculture production comparative value[i]

region	Agriculture production comparative advantage	region	Agriculture production comparative advantage	region	Agriculture production comparative advantage
shanghai	-1.0332	liaoning	-0.1035	hebei	0.2003
zhejiang	-0.6481	hunan	-0.0454	jiangxi	0.2363
tianjin	-0.601	ningxia	-0.0176	henan	0.3797
guangdong	-0.5538	shanxi	0.0185	neimenggu	0.7265
beijing	-0.548	guizhou	0.0348	xinjiang	0.7315
jiangsu	-0.3447	hubei	0.0742	jilin	0.936
fujian	-0.3145	shaanxi	0.1385	heilongjiang	1.5739
shandong	-0.1468	xizang	0.1443		
yunnan	-0.128	guangxi	0.182		

Note: (1). values in this table from Wanlian Lan (2001) are calculated with data in the year of 1999, in this paper, we suppose that this stability still exist, and agriculture comparative production advantage value in this table shows us a comparative role for research reference, not the current real values.
(2). In the former section, agriculture fiscal support efficiency is calculated with data from 1980~2004, so agriculture comparative production advantage values calculated with data of 1999 are suitable.
It is obvious that, APCA reflects agriculture comparative advantage through production angle, and is a compound index synthesized with 5 sub-index, detail APCA calculation process can be referred as endnote i.

4.2. Comprarative Advantage two: ABVL--Agriculture business value of location

ABVL reflects competitiveness of variable industry during regional trade; it's an index measuring comparative advantage: if the value is over zero, it denotes that the industry can satisfy the regional requirement and has some surplus to occupy outside regional market, and is considered of comparative advantage. If the value is less than zero, is regarded as comparative disadvantage. detail ABVL calculation formula can be referred as endnote ii. We rank ABVL by increasing order in table 4.

Table 3. ABVL--Agriculture business value of location[ii]

region	**Agriculture business value of location**	region	**Agriculture business value of location**	region	**Agriculture business value of location**
Shanghai	-0.88	Shandong	-0.04	He'nan	0.47
Beijing	-0.77	Fujian	0.05	Yun'nan	0.48
Tianjin	-0.71	Hubei	0.05	Jiangxi	0.55
Guangdong	-0.34	Shaan'xi	0.08	Neimeng	0.59
Shanxi	-0.32	Hebei	0.11	Guizhou	0.75
Zhejiang	-0.32	Ningxia	0.15	Guangxi	0.76
Liaoling	-0.24	Hu'nan	0.43	Xi'zang	0.87
Jiangsu	-0.24	Xinjiang	0.44		
Heilongjiang	-0.2	Jilin	0.46		

Note: (1). values in this table from Lang Wei (2007), and are mean value calculated with data from 1999 to 2003.
(2). In the former section, agriculture fiscal support efficiency is calculated with data from 1980~2004, also agriculture business values of location calculated with data of 1999 to 2003 are suitable.

5. K-means Clustering

In statistics and machine learning, K-means clustering is an unsupervised learning algorithm that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean [10]. It is similar to theexpectation-maximization algorithm for mixtures of Gaussians in that they both attempt to find the centers of natural cluster in the data. Formally, given a set of sample data $x_1, x_2, ... x_n$ where $x_i \ (i \in [1, n])$ is a d-dimensional real-valued vectors, then K-means cluster aims to partition this set into k partitions $S = \{S_1, S_2, ... S_k\}, k < n$ so as to minimize the within-cluster sum of squares as follows:

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{xj \in Si} \left\| xj - \mu i \right\|^2 \qquad (6)$$

Where μi is the mean of Si . In general the K-means clustering problem is NP-hard, so a variety of heuristic algorithms are generally used. The most common algorithm uses an iterative refinement technique. Given an initial set of k means $m_1^{(1)},...,m_k^{(1)}$, which man be specified randomly or by some heuristic, the algorithm proceeds by alternating between two steps: (a)Assignment step: assign each sample data into the cluster with the closet mean.

$$S_i^{(t)} = \{xj : \left\| xj - m_i^{(t)} \right\| \le \left\| xj - m_{i*}^{(t)} \right\| \ \ for\ i^* = 1,...,k\} \qquad (7)$$

(b) Update step: calculate the new means to be the centroid of the sample data in the cluster.

$$m_i^{(t+1)} = \frac{1}{\left| S_i^{(t)} \right|} \sum_{xj \in S_i^{(t)}} xj \qquad (8)$$

Above algorithm is deemed to have converged when the assignments no longer change.

6. Clusting Analysis

6.1. Discussion

Through analysis in former sections, now for each province in China, there are 3 agricultural values in this paper: agriculture fiscal support efficiency (AFSE— β_i in table 1), agriculture comparative advantage one (APCA value in table 2) and agriculture comparative advantage two (ABVL value in table 3),

From table 1 it is obvious that the agriculture fiscal support efficiency is different, β_i values are from 0.745 (Guangdong) to 1.423 (Xinjiang). It tells us that some regions have comparative efficiency than other regions. As we all know, agriculture is greatly supported by fiscal for almost all countries for its natural weakness, farmer income improvement and important strategy status for national stability. While inside a large country (like China) almost all regions support agriculture with fiscal for certain reasons, although resulting in different efficiency for different region.

In economy theory, comparative advantage is more while at the same time, different region has different agricultural comparative advantage, for economy development, comparative advantage is always the primary consideration, so to local agriculture development, efficiency vs. advantage, which is more important?

6.2. Single Clustering

In this subsection clustering method is used to analyze efficiency and advantage together. Efficiency values are shown in table I as β_i **,** and advantage values are listed in table II.

From table I and table II, we cannot conclude directly from these values that what is the comparative efficiency or comparative advantage of specific province is. So k-means clustering method is used to classify advantage and efficiency respectively into 3 clusters as

H, M and L (H for high advantage or high efficiency, M for medium advantage or medium efficiency, L for less advantage or less efficiency).

- Comparative advantage provinces: H-A provinces:

Provinces in this cluster all have high comparative advantage, including: Neimeng, Jilin, He'nan, Guangxi, Guizhou, Xinjiang, Hu'nan, Yun'nan, Jiangxi, Xizang;.

- Comparative efficiency provinces: H-E provinces:

Provinces in this cluster all have high comparative efficiency, including: Tianjin, Neimeng, Liaoling, Jinin, Fujian, He’nan, Guangxi, Guizhou, Xinjiang;

From above single-type clustering of all provinces, some provinces have comparative advantage, and some provinces have comparative efficiency, for different province respectively belongs to H-A cluster or H-E cluster, such as Xizang in H-A cluster, while Liaoling in H-E cluster, so which one is more suitable for agriculture development? The conclusion is fuzzy from single-type clustering. So 2-dimensional k-means clustering method is used, which combines efficiency and advantage together to classify.

6.3. Two Dimensional Clustering

From single-type clustering of all provinces, some provinces have comparative advantage, and some provinces have comparative efficiency. If we form a combination for all provinces, there are nine combinations at most. For advantage and efficiency respectively has three sorts as H, M and L, so each province has a advantage-efficiency combination, for example Beijing has (L,L) combination while Xinjiang has (H,H) combination.

In this paper, comparative advantage values from table 3 are adopted for analysis. From Table 3, we can find that for regions as Beijing, Shanxi, Zhejiang and Guangdong all have (L, L) advantage-efficiency combinations, to this kind of region, it has absolute agriculture development disadvantage from both aspects; regions as Neimeng, Jilin, He’nan, Guangxi, Guizhou and Xinjiang all have (H, H) advantage-efficiency combinations. To these regions they have absolute agriculture development advantage from both sides. While To regions other than these two kinds, it is hard for them to be decided belonging to which type of regions, whether suitable for agriculture development or not. So hybrid information clustering, that is 2-dimensional k-means clustering method, is secondly used. The clustering results are shown as figure 1.

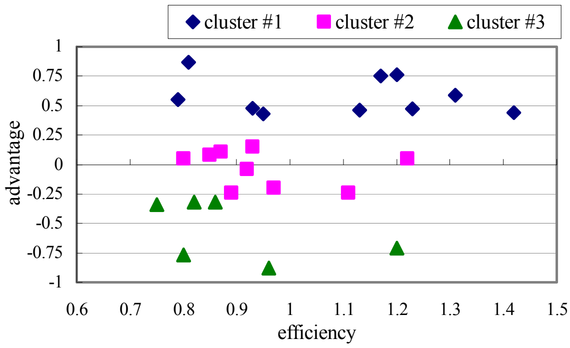

Figure 1. Clustering results

For details, each cluster includes following provinces:

- Cluster #1:High suitable for Agriculture development

Neimeng(H,H), Jilin(H,H), He'nan(H,H), Guangxi(H,H), Guizhou(H,H), Xinjiang(H,H), Hu'nan(H,M), Yun'nan(H,M), Fujian(M,H), Liaoning(M,H);

- Cluster #2: Medium suitable for Agriculture Development

Heilongjiang(M,M), Jiangsu(M,M), Shandong(M,M), Ningxia(M,M), Jiangxi(H,L), Xizang(H,L), Hubei(M,L), Shaanxi(M,L), Hebei(M,L);

- Cluster #3: Least suitable for Agriculture Development

Beijing (L,L), Shanxi (L,L), Zhejiang(L,L), Guangdong(L,L), Tianjin(L,H), Shanghai (L,M)

Now we can get some clear and useful conclusions from the clustering results:

- As a whole, when comparative advantage and comparative efficiency are combined together, provinces in cluster #1 are more suitable for agriculture development than those in cluster #2, and those in cluster #2 are better than those in cluster #3.
- For the hybrid information clustering, comparative advantage plays decisive role, most provinces with H type advantage are classified in cluster #1. But efficiency also has important influence on advantage, for two provinces as Jiangxi and Xizang, both have H advantage and L efficiency, they are classified into cluster #2, maybe too low efficiency (jiangxi as 0.792, Xiang as 0.811) decides these two provinces are medium suitable for agriculture development.

7. Conclusion

From the above clustering results, combined with agriculture advantage and agriculture fiscal support efficiency 25 provinces in China are classified into 3 clusters by 2-dimensional k-means clustering method. It is obvious that different clusters should be developed with different strategy.

In this paper, it is proposed that both comparative efficiency and comparative advantage should be considered based on quantitative analysis for agriculture economy. Comparative advantage is important while efficiency should also not be ignored, to above three clusters, 3 kinds of development strategies are proposed as follows:

- Cluster #1—preference development strategy:

Provinces in this cluster all have high comparative advantage, agriculture preference development strategy is proposed to these regions, and more fiscal support should be given to these areas intentionally.

- Cluster # 2--:maintaining development strategy:

For provinces in this cluster all have M type comparative advantage and H, M or L type efficiency, agriculture maintaining development strategy is proposed for their agriculture resource;

- Cluster #3 edge development strategy :

For provinces in this cluster, all have agriculture development disadvantage and L or M type efficiency, so agriculture edge development strategy is proposed.

References

[1] Li Huanzhang, Qian Zhonghao. Agriculture fiscal support policy and China agriculture increase: causality and structure analysis [J]。 China countryside economy, 2004 (8).
[2] Shen Xiaoli, An Longsong. Experience study of relation between Hu'nan agriculture fiscal support and its agriculture economy increase [J]. Journal of Hu'nan Agriculture University, 2007 (2).
[3] Wei Lang. Study of influence of agriculture fiscal support on China western region agriculture economy increase [J]. Financial Science, 2006 (4).
[4] Wei Lang. Study of influence of agriculture fiscal support on China agriculture economy increase [J]. Journal of Central University of Finance and Economy , 2007 (9).
[5] Lan Wanlian. Agriculture Comparative Advantage of Each China Province and Analysis of Agriculture Relative Proportion. Agriculture Technology Economy.2001(2).
[6] Yu Zhanping. Agriculture Comparative Advantage Theory in New Stage and Empirical Research about Tianjin. China Rural Economy. 2003(9).
[7] Huang Zhaoying. Analyze transition from agriculture comparative advantage to competitive advantage. Agriculture Modernization Study.2006(3).
[8] Justin Yifu Lin & Fang Cai & Zhou Li, 1994. "China's economic reforms : pointers for other economies in transition?," Policy Research Working Paper Series 1310, The World Bank.
[9] Justin Lifu Lin, 2004. "Development strategy, Transition and Challenges of Development in Lagging regions," Development Economics Working Papers 447, East Asian Bureau of Economic Research..
[10] Justin Lifu Lin, 2004. "Development Strategies for Inclusive Growth in Developing Asia," Development Economics Working Papers 444, East Asian Bureau of Economic Research.
[11] Lin, Justin Yifu, 2003. "Development Strategy, Viability, and Economic Convergence," Economic Development and Cultural Change, University of Chicago Press, vol. 51(2), pages 276-308.
[12] E. P. Xing, A. Y. Ng, M. I. Jordan, S. Russell. Distance Metric Learning with Application to Clustering with Side-information. Neural Information Processing Systems, Vol. 15, pp.505-512, 2003.

---

[i] The method for calculating agriculture production comparative advantage refers to Wanlian Lan (2001). He selected five index to analyze agriculture production comparative advantage as following:

(1). Factor endowment index: C1= average household plowland area/capital per person of industry enterprise; this index reflect factor endowment condition for agriculture production;

(2). Agriculture capital comparative index: C4=farmer's household agriculture fixed capital/ farmer's household production total capital, this index reflect agricultural comparative condition about physical capital.

(3). Productivity: C2=agriculture labor productivity/non-agriculture industry labor productivity, this index reflect agriculture productivity condition.

(4). Agriculture opportunity cost: C3= farmer household operation income/farmer labor reward, this index try to reflect farmer's opportunity cost for agriculture production.

(5) agriculture comparative profit index: C5= revenues from agriculture/ total fiscal income, this index reflects agricultural comparative economy interest.

Then above 5 index produce one compound comprehensive index as following manner:

$$C = \sum_{i=1}^{5} W_i C_{ji}$$

In which, C_{ji} represents C_i value of j province; while W_i is weight for each index, and after refer to opinions from many experts, W_1 to W_5 respectively values 0.30, 0.15, 0.15, 0.25 and 0.15.

[ii] The method for calculating agriculture business value of location is as follows:

$$LQ_{ij} = \frac{L_{ij}}{\sum_j L_{ij}} \Bigg/ \frac{\sum_j L_{ij}}{\sum_i \sum_j L_{ij}} - 1$$

In which, L_{ij} denotes adding value of j industry of I region, LQ_{ij} reflects competitiveness of variable industry joining regional trade, it's a index measuring comparative advantage: if the value is over zero, it denotes that the industry can satisfy the regional requirement and has some surplus to occupy outside regional market, and is considered with comparative advantage. If the value is less than zero, is regarded as comparative disadvantage.