

Component Data Combination Forecasting Model and Its Application in Prediction and Analysis of Industrial Structure

Lingrong Zhao*, Ying Ping and Yuanhong Luo

Shanghai Ocean University, School of Economics and Management, Shanghai, 201306, China

*Corresponding author

Abstract—Composition data represent the relative information of data, not absolute information. The determination and restriction make the analysis of composition data different from other general data, and need to be transformed and re-analyzed first. For a series of chronological data collected, the paper uses two components of data formed by asymmetric logratio transformation and spherical coordinate transformation to establish the forecasting model respectively, and the two forecasting models are combined and forecasted to establish the forecasting model. According to the forecasting method, a forecasting model of Chinese fishery industrial structure is established and the future industrial structure of Chinese fishery is predicted.

Keywords—logratio transformation; spherical coordinate transformation; combined forecasting model; industrial structure

I. INTRODUCTION

Nowadays, component data is widely used in data analysis in many fields such as society, economy, technology and so on. It is usually used to study the ratio of various parts as a whole. For example, to study the structure of three industries in China, we can use the first The proportion of industries, secondary industries and tertiary industries as a percentage of the gross national product (GNP) is expressed, but the proportion has its own uniqueness. The sum of these proportions is equal to one. Expressed in mathematical expressions, the component data refers to any non-negative p-dimensional vector $X = (x_1, x_2, \dots, x_p)$, the value of p components of X is greater than or equal to zero and less than or equal to one.

In forecasting, due to the particularity of the composition data, the forecast of composition data must be transformed first, and then the transformed data should be predicted, and finally the predicted data should be inversely solved. There are also many single-predictive models in the forecasting of component data, and these models will show better results in different aspects. In order to improve the accuracy of forecasting, the combined forecasting model uses different forecasting models for forecasting the same thing. By comprehensively using the information provided by each forecasting model, we can get better forecasting results. Generally, the combination forecasting method uses n ($n \geq 2$) different single forecasting models to forecast at first, and combines a plurality of single forecasting models through a certain criterion to form a combined model, and then uses the combined model to carry out the combined forecasting calculation.

For a series of chronological data collection, the paper

proposes to use non-symmetric logratio transform and spherical coordinate transform to form the component data, to do non-linear dimension reduction on the component data to obtain the vector data which is consistent with the degree of freedom of the component data, and to establish composition data prediction modeling. The article also uses the proposed composition forecast model to predict the future development trend of Chinese fishery industry structure.

II. COMPOSITION DATA COMPOSITION PREDICTION MODEL MODELING METHOD

According to the data of asymmetric logratio transformation and spherical coordinate transformation, the regression analysis is used to obtain two sets of prediction results. Based on the minimum weighted optimal Aitchison distance sum between the predicted value and the true value, the combined forecasting model gives better predictions. The basic steps are as follows:

A. Model 1: Asymmetric Data Logratio Transform

Let $x = (x_1, x_2, \dots, x_D)$ be the component vector, so that:

$$y_D = \ln x_D / x_i, D=1, 2, \dots, i-D$$

The above transformation is called asymmetric logratio transformation, which can overcome the "set and limit" of the component data, partially to eliminate the complete correlation between the components in order to use the least square method to solve. If y_i is used as x_i , here is the following expression.

$$\begin{cases} x_i = e^{y_i} / \left(1 + \sum_{j=1}^{D-1} e^{y_j} \right) \\ x_D = 1 / \left(1 + \sum_{j=1}^{D-1} e^{y_j} \right) \end{cases}$$

B. Model 2: Data Transformation of the Spherical Coordinates

Ball coordinate transformation is proposed by Wang Huiwen in 2004, and it has been applied to the forecast. From the form of spherical coordinate transformation, it can solve the predicament of zero component transformation in logratio transformation.

Non-linear transformation of the original data: $y_i(t) =$

$\sqrt{x_i(t)}, i = 1, 2, \dots, T$, Let

$Y(t) = (y_1(t), y_2(t), \dots, y_D(t)) t = 1, 2, \dots, T$. So,

$$\|Y(t)\|^2 = \sum_{i=1}^D y_i(t) = 1$$

As can be seen from the above formula, the data is spread over a D-dimensional hypersphere with a radius of 1. Transform the data from the Cartesian coordinate system to the spherical coordinate system yields:

$$y_1(t) = \sin\theta_2(t)\sin\theta_3(t)\dots\sin\theta_D(t)$$

$$y_2(t) = \cos\theta_2(t)\sin\theta_3(t)\dots\sin\theta_D(t)$$

⋮

$$y_D(t) = \cos\theta_D(t)$$

By the above inverse transformation:

$$\begin{cases} \theta_D = \arccos y_D \\ \theta_{D-1} = \arccos\left(\frac{y_{D-1}}{\sin\theta_D}\right) \\ \theta_{D-2} = \arccos\left(\frac{y_{D-2}}{\sin\theta_D \sin\theta_{D-1}}\right) \\ \vdots \\ \theta_2 = \arccos\left(\frac{y_2}{\sin\theta_D \sin\theta_{D-1} \dots \sin\theta_3}\right) \end{cases}$$

C. Sum of Square Error and Weight Method

In this paper, the weights are mainly based on the sum of squares of errors, however, for the composition data, because the composition data is a vector, the error between the predicted value and the true value can not simply be poor, but the distance between the component data and the vector is regarded as the error between the predicted value and the true value, that is to say, the weight coefficient is obtained based on the sum of squares of the Aitchison distances between the predicted value and the true value.

Aitchison distance: Let two component data vectors: $x = (x_1, x_2, \dots, x_D)$, $y = (y_1, y_2, \dots, y_D)$, and then the Aitchison distance is:

$$d_a(x, y) = \left(\sum_{i=1}^D \left(\ln \frac{x_i}{g(x)} - \ln \frac{y_i}{g(y)}\right)^2\right)^{1/2}$$

$$g(x) = (x_1 x_2 \dots x_D)^{1/D}, g(y) = (y_1 y_2 \dots y_D)^{1/D}$$

The following steps for weight coefficient:

(1) Assuming that the component vector $x^t = (x_1^t, x_2^t, \dots, x_D^t)$ has n ($n \geq 2$) kinds of prediction models and x^t is the true value of time t, the i th prediction model is first obtained at the tth time when the predicted value is $y_i^t = (y_1^t, y_2^t, \dots, y_D^t)$ and the prediction error is $d_n(x^t, y_i^t)$, then the n kinds of single prediction models are combined, that is:

$$\hat{x}^t = \sum_{i=1}^n \beta_i y_i^t, t = 1, 2, \dots, N; i = 1, 2, \dots, n$$

Where i denotes the weight of the ith predictive model in the combined forecast, and is satisfied $\sum_{i=1}^n \beta_i = 1$.

(2) Let SSc be the sum of Aitchison distances of the combined forecasting model, that is:

$$SS_c = \sum_{t=1}^N d_a^2(x^t, \hat{x}^t) = \sum_{t=1}^N \sum_{i=1}^n \sum_{j=1}^n \beta_i \beta_j d_a(x^t, y_i^t) d_a(x^t, y_j^t)$$

Weights β can be obtained by minimizing SSc using Lagrange multiplier method under constraints $\sum_{i=1}^n \beta_i = 1$.

III. EMPIRICAL RESEARCH

According to the trend of the three industrial structures and the proportion of in the fishery industry, the structure of China's fishery industry has been continuously optimized: the primary industry has been declining, and the proportion of the secondary and tertiary industry has been on the rise. Between 1999 and 2016, the proportion of primary fishery in China dropped from 64% to 51%, the proportion of secondary industry remained at 25%, the proportion of tertiary industry rose from 11% to 26%, and the tertiary industry, which is mainly based on fishery circulation industries and services, is constantly being valued by the state. The structure of Chinese fishery industry is going through pyramid structure, dumbbell structure and inverted pyramid structure in sequence. As shown in Figure 1, the proportion of primary industry, secondary industry and tertiary industry in fishery in 1994 is 64:25:11, that is, the pyramid structure. The proportion of the three fisheries industries in 2004 was 57:22:21 and in 2016 it was adjusted to 51:23:26. The comparison shows that the proportion of China's primary fishery industry is declining while the proportion of the tertiary industry, mainly fishery circulation and service industry, is rapidly increasing.

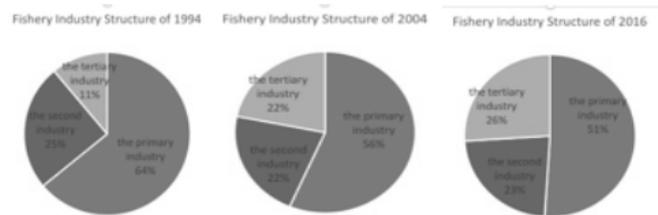


FIGURE I. DISTRIBUTION OF CHINESE FISHERIES INDUSTRIAL STRUCTURE IN 1994, 2004 AND 2016

According to the calculation results of the model, in recent years, the development trend of Chinese fishery is still that the proportion of the primary industry is gradually reduced and the proportion of the second and tertiary industries continue to rise: In 2017, the proportion of the primary fishery industry will reduce to 48%, below 50% for the first time, showing a declining trend in the next five years. With the development of economy and the construction of a well-off society in an all-round way, the tertiary industry with circulation and service industries will

develop rapidly, accounting for more than the proportion of the second industry with the processing industry being the main secondary industry, so that Chinese fishery industry structure shows a "one two two" distribution pattern. Overall, in the next five years, the proportion of secondary and tertiary industries in fisheries of China will continue to increase, reaching 54% by 2021, and the structure of the fishery industry will continue to be rationalized.

A. Data Preprocessing

In order to maintain a healthy and sustainable fishery industry, China established the China Fisheries Mutual Insurance Association in 1994. In the same year, the central authorities emphasized that Chinese fishery industry is an

important industry for rural economy development. It can increase peasants' income (fishing) and adjust the rural industrial structure. To this end, we have selected the data of the total output value of fishery in our country from 1994 to 2016 for a total of 23 years. The data come from the "China Fisheries Statistical Yearbook" from 1995 to 2017.

B. Model Building

First of all, for the proportion of the three major industries in fishery, respectively, do asymmetric logratio transform and transformation of the ball coordinates in Table 1.

TABLE I. CHINESE FISHERY TERTIARY INDUSTRY STRUCTURE DATA AND THE TRANSFORMED DATA

Year	x1	x2	x3	asymmetric logratio transformation		Spherical coordinates transformation	
				y1	y2	θ_2	θ_3
1994	0.64	0.25	0.11	1.7385	0.8021	1.0114	1.2294
1995	0.65	0.25	0.10	1.8543	0.9049	1.0143	1.2467
1996	0.66	0.26	0.09	2.0177	1.0736	1.0131	1.2707
1997	0.64	0.25	0.11	1.7520	0.8267	1.0089	1.2320
1998	0.64	0.25	0.11	1.7128	0.7665	1.0136	1.2248
1999	0.58	0.27	0.15	1.3906	0.6230	0.9728	1.1797
2000	0.54	0.28	0.18	1.1197	0.4663	0.9459	1.1369
2001	0.47	0.30	0.23	0.6931	0.2480	0.8958	1.0664
2002	0.41	0.33	0.26	0.4615	0.2645	0.8346	1.0390
2003	0.58	0.22	0.21	1.0241	0.0556	1.0186	1.0990
2004	0.57	0.22	0.22	0.9655	0.0103	1.0156	1.0878
2005	0.55	0.23	0.23	0.8866	-0.0039	1.0010	1.0753
2006	0.53	0.24	0.23	0.8413	0.0350	0.9817	1.0711
2007	0.52	0.24	0.24	0.7907	0.0381	0.9693	1.0639
2008	0.53	0.25	0.22	0.8689	0.1010	0.9728	1.0794
2009	0.52	0.23	0.25	0.7417	-0.0539	0.9792	1.0505
2010	0.52	0.24	0.24	0.7820	0.0000	0.9761	1.0601
2011	0.53	0.24	0.24	0.7855	-0.0190	0.9813	1.0594
2012	0.52	0.24	0.24	0.7805	-0.0045	0.9768	1.0596
2013	0.52	0.23	0.24	0.7599	-0.0443	0.9813	1.0539
2014	0.52	0.23	0.25	0.7516	-0.0494	0.9805	1.0523
2015	0.52	0.23	0.25	0.7247	-0.0933	0.9800	1.0425
2016	0.51	0.23	0.26	0.6737	-0.1226	0.9789	1.0413

Table 1, x1, x2, x3, respectively, the proportion of the first, second and third industries, asymmetric logratio transformed data y1, y2, spherical coordinates transformed data θ_2 , θ_3 .

Based on the data in Table 1, regression prediction of y1, y2, θ_2 and θ_3 respectively is performed, and then inversely they are transformed into the component data to obtain the fitting value of the tertiary industry structure of Chinese fishery. Here's a fitting function for regression prediction using stata:

Fitting Function of Asymmetric Logration Transformation:

$$y1 = -0.053t + 1.681$$

$$y2 = -0.05t + 0.852$$

Fitting function of spherical coordinate transformation:

$$\theta_2 = -0.00065t + 0.986$$

$$\theta_3 = -0.0097t + 1.226$$

Based on the theory of combined forecasting, the sum of squared error of the combined forecast is obtained when $w = (0.25, 0.75)$. So get the regression predictive fit values based on the asymmetric logratio transform, the regression predictive fit values based on the spherical coordinate transform, and the

combined predictive fit values of the two methods.

According to the calculation results of the model, in recent years, the development trend of Chinese fishery is still that the proportion of the primary industry is gradually reduced and the proportion of the second and tertiary industries continue to rise: In 2017, the proportion of the primary fishery industry will reduce to 48%, below 50% for the first time, showing a declining trend in the next five years. With the development of economy and the construction of a well-off society in an all-round way, the tertiary industry with circulation and service industries will develop rapidly, accounting for more than the proportion of the second industry with the processing industry being the main secondary industry, so that Chinese fishery industry structure shows a "one two two" distribution pattern. Overall, in the next five years, the proportion of secondary and tertiary industries in fisheries of China will continue to increase, reaching 54% by 2021, and the structure of the fishery industry will continue to be rationalized.

As can be seen from Figure 1, 2 and 3, the fitting effect of combined forecasting is better than that of asymmetric variant forecasting and spherical coordinate forecasting, and the Aichision distance is smaller, which means the combined forecast can more accurately reflect the future development trend.

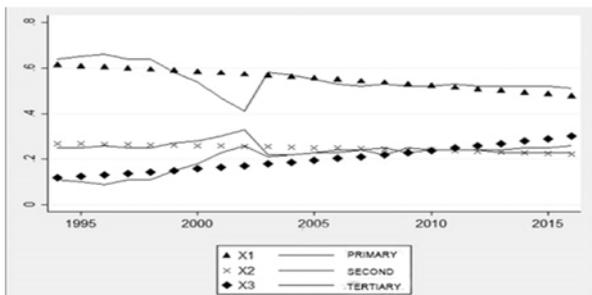


FIGURE II. ASYMMETRIC TRANSFORMATION REGRESSION PREDICTION

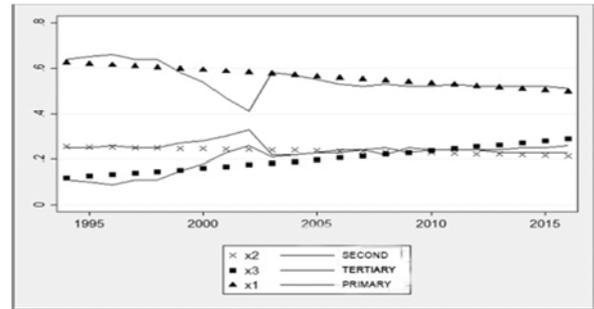


FIGURE III. ASYMMETRIC TRANSFORMATION REGRESSION PREDICTION

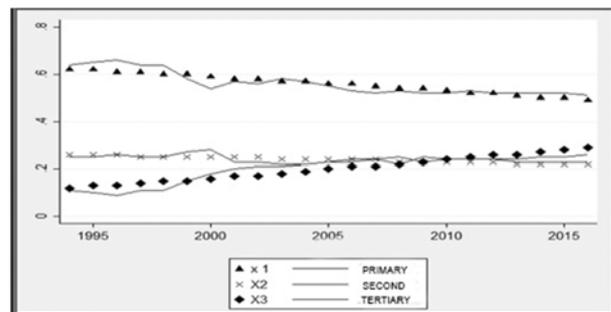


FIGURE IV. COMBINATION REGRESSION PREDICTION

C. Model Application

The above three models are respectively used to predict the structure of Chinese fishery industry in the

next five years. It is assumed that the indicators will maintain the current growth rate in the next five years. The forecast results are shown in Table 3.

TABLE II. 2017-2021 FORECAST RESULT

Year	Asymmetric logratio transformation regression prediction			Spherical coordinates transformation regression prediction			Combination regression prediction		
	x1	x2	x3	x1	x2	x3	x1	x2	x3
2017	0.4687	0.2199	0.3114	0.4898	0.2121	0.2981	0.4845	0.2140	0.3014
2018	0.4606	0.2167	0.3227	0.4832	0.2097	0.3070	0.4776	0.2115	0.3109
2019	0.4524	0.2135	0.3341	0.4766	0.2074	0.3160	0.4706	0.2089	0.3206
2020	0.4440	0.2102	0.3458	0.4700	0.2050	0.3251	0.4635	0.2063	0.3303
2021	0.4356	0.2068	0.3577	0.4633	0.2025	0.3342	0.4563	0.2036	0.3401

According to the calculation results of the model, in recent years, the development trend of Chinese fishery is still that the proportion of the primary industry is gradually reduced and the proportion of the second and tertiary industries continue to rise: In 2017, the proportion of the primary fishery industry will reduce to 48%, below 50% for the first time, showing a declining trend in the next five years. With the development of economy and the construction of a well-off society in an all-round way, the tertiary industry with circulation and service industries will

develop rapidly, accounting for more than the proportion of the second industry with the processing industry being the main secondary industry, so that Chinese fishery industry structure shows a "one two two" distribution pattern. Overall, in the next five years, the proportion of secondary and tertiary industries in fisheries of China will continue to increase, reaching 54% by 2021, and the structure of the fishery industry will continue to be rationalized.

IV. THE CONCLUSION AND INSPIRATION

Industrial structure development and economic development are a simultaneous process of adaptation. The level of economic development in a country is directly related to the high level of industrial structure. The structure of Chinese fishery industry is in line with this law of development and its transition from a lower level to a higher level is being gradually realized. In this paper, composition forecasting model is applied to further study the future development trend of Chinese fishery industry structure. The empirical results are as follows: From the perspective of the future development of fishery industry structure, the proportion of Chinese fishery secondary industry and the primary industry continues to increase. The proportion of secondary and tertiary industries is on the rise. By 2020, the proportion of secondary and tertiary industries in China's fisheries will reach 53%, reaching the output level of secondary and tertiary industries in fisheries in 2020 set by the Ministry of Agriculture in the 13th Five-Year Plan for National Fisheries Development—the proportion of more than 50% of the development goal.

In order to promote Chinese successful completion of its development goals, our policy recommendations are to focus on the specialization of the fishery industry, extend the industrial chain, actively develop the tertiary industry, promote the development of recreational fisheries, actively train new rural management entities and promote coordination among the integrate development of three major industries and speed up economic restructuring. The economy mainly relying on investment in the past should be shifted to one relying on consumption and investment to promote a better and faster development of Chinese fishery economy.

Based on the proposed forecasting model of composition data, the paper establishes the forecast model of the development of Chinese fishery industry

structure. Through empirical analysis, it is concluded that the composition forecast data model can correctly reflect the data movement rules, predict and analyze the data reasonably and effectively.

ACKNOWLEDGEMENTS

The article is supported by Fishery Management Project 2016 (Q700) and Luo Zhaoyao Fund of Shanghai Ocean University.

REFERENCE

- [1] Xue Bai. The transformation of economic growth mode based on the optimization of industrial structure—function mechanism and measurement[J].*Management Science*.2009,10:112-120
- [2] Wang Yanqi. Multi-plan analysis of agricultural structure optimization in Heilongjiang Province [J]. *Chinese Agricultural Resources and Regional Planning*. 2008, 10:18-21
- [3] Yang Zhongna, Tang Jijun. The influence of agricultural structure change on agricultural economic growth in South Xinjiang[J]. *Chinese Agricultural Resources and Regional Planning*. 2014, 6:85-92
- [4] Zhu Xiaohua, Deng Baoyi. Empirical analysis of the impact of China's industrial structure on economic growth [J]. *Business Economy*. 2013, 7: 132-136
- [5] Yang Lin. Research on the adjustment of China's fishery industry structure under the constraints of resources and environment [J]. *Rural economy*. 2004,8:28-31
- [6] Yan Fangfang, Ping Yu. Empirical Study on the Relationship between Consumption Demand Structure and Industrial Structure: Taking China's Fisheries as an Example [J]. *Chinese Agricultural*
- [7] YU Fangfang, YANG Ningsheng. Optimization of China's fishery industry structure after the WTO era based on grey relational model[J]. 2010, 6:614-616
- [8] Li Daliang, Shi Lei. Research on the optimization of fishery industry structure in China [J]. *Chinese fishery economy*. 2009, 4:41-45