

# Construction of Prediction Model for Online Shopping Expenditure of Rural Residents in China

Can Wang

School of Mathematics  
Chengdu Normal University  
Chengdu, China

**Abstract**—The development of e-commerce in rural areas is not only an important part of rural economy, but also closely related to the transformation of rural residents' consumption and life mode. Regarding to the issue of online shopping expenditure for rural residents in China, the paper has successively set up a quadratic curve exponential smoothing model, Grey GM (1, 1) model and combination prediction model based on the two, and applied a new method to determine the combined weight. Based on the data of online shopping expenditure of rural residents in China from 2010 to 2017, the three models are calculated and tested by using relevant software such as SPSS, and the results show that the prediction data of the combination prediction model is more stable. Finally, the paper forecasts and analyzes the data of online shopping expenditure of rural residents in China in the next three years and puts forward the control suggestions.

**Keywords**—online shopping of rural residents; exponential smoothing model; grey system; combination prediction

## I. INTRODUCTION

In the era of highly developed informatization, the popularity of electronic products supporting e-commerce is relatively high, and shopping has become an important part of the people's life. However, in rural areas, the influence of the Internet is relatively low, and the proportion of shopping on the Internet is also relatively low. Therefore, rural residents group is an important reserve resource for the further development of online shopping market. The mining of its relevant data is conducive to the further prosperity of China's economic market, and at the same time, it can also bring convenience to the life of rural residents, so that they can enjoy the joy of online shopping, which can benefit both parties. However, there are some exceptions, for example, some people don't know the process of online shopping, and they don't know how to do it if they want shopping online; and some people feel insecure about online shopping, worried about disclosing their personal information and bank card passwords; and some people feel worried that the goods bought online are quite different from the physical objects, such factors will undoubtedly affect the enthusiasm of rural residents to buy online.

Most market signals suggest that growth in Chinese e-commerce spending will slow down in the next few years. It remains to be seen whether this trend is a temporary

phenomenon or a sign that China's e-commerce industry is maturing. At present, the research on this part mainly focuses on the sampling survey of the current situation of rural online shopping in China [5], or the promotion of rural e-commerce to local economy [6]- [8], etc. And the development of statistical prediction model has been relatively mature and widely used. For example, XuYaqing et al. built the prediction model [1] of agricultural product price, and XuMinjie et al. constructed the model of network public opinion data combination prediction[2], which both achieved good results.

Based on the data mining method, this paper takes into account the factors such as economic and business model changes to establish the prediction model for online shopping expenditure of rural residents in China, and continuously conducts iterative optimization on the model and uses a new combined weight method to construct the combined model, to finally realize the prediction to the trend of the share e-commerce will take in the rural market in the next few years, so as to support the e-commerce to optimize the resource structure and support the government to make relevant policies.

## II. ESTABLISHMENT OF THE PREDICTION MODEL

### A. The Quadratic Curve Index Exponential Smoothing Model

According to the survey data of AliCloud research institute, the online shopping consumption amount of rural residents in China has the trend of curve growth in time series, as shown in figure 1, therefore, this paper firstly establishes the quadratic curve exponential smoothing model, which can consider the growth trend of linear and quadratic parabola at the same time. The calculation steps of the model are as follows:

Firstly, the paper calculated separately the single exponential smoothed value  $S_t^{(1)}$ , the double exponential smoothed value  $S_t^{(2)}$  and the triple exponential smoothed value  $S_t^{(3)}$  during period  $t$ , and

$$\begin{aligned} S_t^{(1)} &= \alpha x_t + (1-\alpha)S_{t-1}^{(1)} \\ S_t^{(2)} &= \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)} \\ S_t^{(3)} &= \alpha S_t^{(2)} + (1-\alpha)S_{t-1}^{(3)} \end{aligned}$$

Then calculated the level value  $A_t$  and the linear increment  $B_t$  during period t, among which

$$\begin{aligned} A_t &= 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \\ B_t &= \frac{\alpha^2}{(1-\alpha)^2} \left[ (6-5\alpha)S_t^{(1)} - (10-8\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)} \right] \end{aligned}$$

Finally calculated the linear increment  $C_t$  during the period t:

$$C_t = \frac{\alpha^2}{(1-\alpha)^2} (S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)})$$

After the period m, i.e. the prediction value  $F_{t+m}$  during the period  $t+m$ :

$$F_{t+m} = A_t + B_t + \frac{1}{2} C_t m^2$$

Among which m is an integer greater than or equal to 1, and  $\alpha$  is the smoothing factor. Although the calculation of the model is a little complicated, it is very effective for the time series prediction which has tendency. The smoothing factor is determined by the SPSS software as  $\alpha=0.6$ .

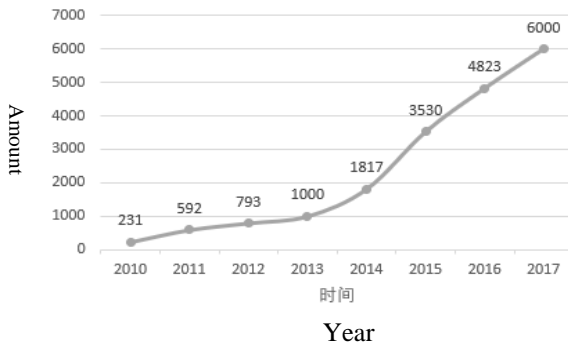


Fig. 1. Time series chart of total amount of online shopping expenditure of rural residents in China (one hundred million yuan/year)

### B. Grey GM(1,1) Model

The grey system itself is a system between the white system and the black system, and the grey prediction obviously involves the uncertain factors. In the forecast of China's rural residents' online shopping expenditure, it

contains uncertain information in the operation of e-commerce, so this paper takes it as the second prediction model.

The prediction model of GM(1,1) is:

$$\hat{X}^{(1)}(k+1) = \left[ X^{(0)}(1) - \frac{\mu}{a} \right] e^{-ak} + \frac{\mu}{a} \quad (k=0,1,2,\dots,n)$$

Solved by the least square method as:

$$\begin{pmatrix} a \\ \mu \end{pmatrix} = (B^T B)^{-1} B^T Y^n$$

Of which

$$B = \begin{bmatrix} -\frac{1}{2} [X^{(1)}(1) + X^{(1)}(2)] & 1 \\ -\frac{1}{2} [X^{(1)}(2) + X^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2} [X^{(1)}(n-1) + X^{(1)}(n)] & 1 \end{bmatrix}$$

$$Y_n = \begin{bmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \vdots \\ X^{(0)}(n) \end{bmatrix}$$

$X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n)\}$  is the initial time series and corresponding observed values,

$X^{(1)} = \{X^{(1)}(1), X^{(1)}(2), \dots, X^{(1)}(n)\}$  is a new series generated after the initial series is accumulated.

### C. Combination Prediction

Each prediction model has its own limitations. In order to minimize the errors, the combination prediction is a good method, but the biggest problem of combination prediction is the determination of weight. At present, the relatively common weight value determination methods are analytic hierarchy process (AHP), error analysis method, least square method, and optimal combination method, etc. In this paper, the combined weights [3] are reconfirmed on the basis of the mean absolute percentage error (MAPE) criterion and the AHP method, that is, the weight of the  $i^{\text{th}}$  model in the combined model is:

$$\lambda_i = \frac{a_i b_i}{\sum_{i=1}^n a_i b_i} \quad (1)$$

Of which  $a_i$  is the coefficient determined by the  $i$ th model based on the MAPE criterion and

$$a_i = \frac{1}{MAPE_i} \bigg/ \sum_{i=1}^n \frac{1}{MAPE_i} \quad (2)$$

$b_i$  is the coefficient[4] determined by the  $i$ th model based on the AHP method, and the AHP method mainly consists of three steps of constructing the judgment matrix A, determining weight and checking consistency. In this paper, there are two single models which are second-order judgment matrix and may be relieved of the consistency check. The determination of each element of the judgment matrix A is shown in "Table I".

TABLE I. THE DETERMINATION OF EACH ELEMENT OF THE JUDGMENT MATRIX

Matrix Element $a_{ij}$	The Comparison of the Two Goals
1	Equally important
3	A little more important than
5	Obviously more important than
7	Much more important than
9	very much more important than
2, 4, 6, 8	Between the above two adjacent situations
The reciprocal of each of the above numbers	Compare the two goals the other way around

Then the weight is determined by the following steps:

Normalize each column of the matrix:

$$\bar{a}_{ij} = \frac{a_{ij}}{\sum_{i=1}^n a_{kj}} \quad i, j = 1, 2, \dots, n$$

Add the rows of the new matrix obtained after normalizing the column:

$$M_i = \sum_{j=1}^n \bar{a}_{ij} \quad i = 1, 2, \dots, n$$

Then normalize the new vector again

$$M = (M_1, M_2, \dots, M_n)^T \quad i = 1, 2, \dots, n$$

$$W_i = \frac{M_i}{\sum_{j=1}^n M_j} \quad i = 1, 2, \dots, n$$

The obtained  $W=(W_1, W_2, \dots, W_n)$  is the corresponding eigenvector, and when  $n$  is equal to 2, it is the required weight.

### III. MODEL APPLICATION AND PREDICTION ANALYSIS

This paper selects the amount of online shopping consumption of rural residents in China from 2010 to 2017 as the research object. The data is from the AliCloud research institute and the China E-business Research Center.

The definition of rural residents involved in the data is the population of permanent rural residents, and online shopping expenditure is the amount of transactions that are sent to the rural areas according to the statistics of the deals made.

#### A. Calculate the Prediction Value of Each Model

##### 1) Quadratic curve exponential smoothing model:

The initial value of the quadratic exponential smoothing method depends on the first two periods of observed values

$x_1$  and  $x_2$  of the original sequence value, therefore, first we

make  $S_1^{(1)} = S_1^{(2)} = S_1^{(3)} = x_1$ , then we

have  $S_2^{(1)} = \alpha x_2 + (1-\alpha)x_1, S_2^{(2)} = \alpha S_2^{(1)} + (1-\alpha)x_1,$

$S_2^{(3)} = \alpha S_2^{(2)} + (1-\alpha)x_1$ , then we make  $m=1$ , the year-by-year prediction values obtained are shown in "Table II".

TABLE II. PREDICTION VALUE OF QUADRATIC CURVE EXPONENTIAL SMOOTHING METHOD AND ITS MEAN RELATIVE ERROR (UNIT: 100 MILLION YUAN)

Year	Actual Value	Level Value	Linear Increment	Parabolic Increment	Prediction Value	Mean Relative Error
2010	231.00	--	--	--	--	--
2011	592.00	546.88	203.06	45.13	--	-
2012	793.00	767.88	113.06	25.13	772.50	2.59%
2013	1000.00	974.13	116.44	25.88	893.50	10.65%
2014	1817.00	1714.88	459.56	102.13	1103.50	39.27%
2015	3530.00	3315.88	963.56	214.13	2225.50	36.95%
2016	4823.00	4661.38	727.31	161.63	4386.50	9.05%
2017	6021.00	5871.25	673.875	149.75	5469.50	9.16%
2018	--	--	--	--	6620.00	--
2019	--	--	--	--	7518.50	--
2020	--	--	--	--	8551.78	--

2) *GM(1,1) model:*

It can be obtained by calculation of the python-related program through the least square method that:

$$\begin{pmatrix} a \\ \mu \end{pmatrix} = \begin{pmatrix} -0.38 \\ 504.16 \end{pmatrix}$$

The obtained prediction model is:

$$\hat{X}^{(1)}(k+1) = 1557.74e^{0.38k} - 1326.74 \quad (k = 0, 1, 2, \dots, n)$$

The model can be used for prediction after inspection, and we can get specific prediction values by the prediction

formula  $X^{(0)}(k+1) = X^{(1)}(k+1) - X^{(1)}(k)$ , as shown in "Table III".

TABLE III. PREDICTION VALUE OF GREY GM(1,1) METHOD AND ITS MEAN RELATIVE ERROR (UNIT: 100 MILLION YUAN)

Year	Actual Value	Time Series Value	Grey Value	Prediction Value	Mean Relative Error
2010	231.00	231.00	231.00	231	0.00%
2011	592.00	823.00	951.12	620.12	4.75%
2012	793.00	1616.00	1571.24	1053.02	32.79%
2013	1000.00	2616.00	2624.26	1039.81	3.98%
2014	1817.00	4433.00	3664.07	2251.65	23.92%
2015	3530.00	7963.00	5915.72	3292.55	6.73%
2016	4823.00	12786.00	9208.27	4814.64	0.17%
2017	6021.00	18807.00	14022.91	6040.38	0.32%
2018	--	--	20063.29	8295.03	--
2019	--	--	28358.32	10054.27	--
2020	--	--	38412.59	15013.62	--

3) *Combined model:*

In case that the quadratic curve exponential smoothing model is model 1 and the GM(1,1) model is model 2, it can be calculated that  $MAPE_1=17.9\%$ ,  $MAPE_2=10.38\%$ ,

and we can obtain  $a_1=0.37$ ,  $a_2=0.63$  when substituting it into the formula (2).

$$A = \begin{pmatrix} 1 & \frac{1}{2} \\ 2 & 1 \end{pmatrix}$$

Construct the judgment matrix according to the weight calculation steps based on the AHP method,

we can obtain  $b_1=0.33$ ,  $b_2=0.67$ , and we can obtain the final weights of the combined model as:  $\lambda_1=0.22$ ,  $\lambda_2=0.78$  when substituting it into the formula (1).

Therefore, the prediction values and errors of the single model and the combined model can be obtained, as shown in Table IV, and the prediction values of online shopping expenditure of rural residents in China in the next three years, as shown in "Table V", can also be obtained through the single model and the combined model.

**TABLE IV. PREDICTION VALUE AND ERROR OF THE SINGLE MODEL AND THE COMBINED MODEL (UNIT: 100 MILLION YUAN)**

Year	Actual Value	Prediction Value and Error of the Model 1	Prediction Value and Error of the Model 2	Prediction Value and Error of the Combined Model
2010	231.00	----	231.00 0.00%	----
2011	592.00	----	620.124.75%	----
2012	793.00	772.50 2.59%	1053.0232.79%	991.31 25.01%
2013	1000.00	893.50 10.65%	1039.813.98%	1007.62 0.76%
2014	1817.00	1103.50 39.27%	2251.6523.92%	1999.06 10.02%
2015	3530.00	2225.50 36.95%	3292.556.73%	3057.80 13.38%
2016	4823.00	4386.50 9.05%	4814.640.17%	4720.45 2.13%
2017	6021.00	5469.50 9.16%	6040.38 0.32%	5914.79 1.76%

**TABLE V. THE FUTURE PREDICTION VALUE OF THE SINGLE MODEL AND THE COMBINED MODEL (UNIT: 100 MILLION YUAN)**

Year	Prediction Value of the Model 1	Prediction Value of the Model 2	Prediction Value of the Combined Model
2018	6620.00	7295.03	7160.02
2019	7518.50	8054.27	7947.12
2020	8551.78	9213.62	9081.25

#### IV. CONCLUSION

Firstly, this paper selected the quadratic curve exponential smoothing model with  $\alpha=0.6$  according to the time series variation trend of online shopping consumption of rural residents in China from 2010 to 2017, then established the GM(1,1) model based on the characteristics of many grey information contained in the electronic business information, and finally established a combination prediction model based on these two models, and applied a new method to determine the coefficient of the combined weight. According to the results obtained, the overall error of the combination prediction model is smaller, the prediction accuracy is the highest, and the data is more stable. Finally, this paper gave the prediction value of online shopping expenditure of rural residents in China in the next three years using the three established models.

Subject to data and model limitations, there are still some improvements that need to be made in this paper:

(1) Sources of data: if there is more official, regular and formal publish of the data of the online shopping expenditure of rural residents in China, the accuracy of the prediction may be higher and more valuable in reference. Of course, this requires relevant organizations and platforms to support and supervise online shopping data.

(2) Data refinement: more of the known data are national data each year. However, according to actual sample surveys, the online shopping is more seasonal, if there is more complete quarterly data, we can choose the prediction models that can reflect the seasonal index, which makes more sense for the guidance of e-commerce development.

(3) Data range: Although this paper has stressed that "the definition of rural residents involved in the data is the population of the permanent rural residents ", however,

because the policy on residence registration has been weakened and the speed of rural urbanization in China has been pushed forward, so there is no unified and clear definition of the data involved in the definition at present. In addition, the definition of "online shopping expenditure is the amount of transactions that are sent to the rural areas among the statistical transaction order", which shows the data in this paper includes the amount of transactions that occurred when the rural migrant workers send the orders directly back to their rural addresses by online shopping. If there is support of more comprehensive and refining data platform, the prediction research in this aspect will be more complete.

The statistics of online shopping consumption expenditure are based on the big data platform. The next step is to combine the intelligent big data platform with the traditional prediction model to further improve the prediction accuracy.

#### REFERENCES

- [1] Xu Yaqing, Wei Yihua, Li Xugang. Construction of Agricultural Commodity Price Prediction Model [J]. Statistics and Decision, 2017 (12): 75-77.
- [2] Xu Minjie, LanYuexin, Liu Bingyue. Research on Prediction Model of Network Public Opinion Data Based on Combination prediction [J]. Intelligence Science, 2016, 34(12): 40-45.
- [3] Li Pei, PengSijun. Application of a New Combined Weight in Combination prediction Model [J/OL]. Journal of Henan University of Science and Technology (Natural Science Edition), 2018(02):87-93.
- [4] Xu Guoxiang, Statistical Prediction and Decision (the Fourth Edition) [M]. Shanghai University of Finance and Economics Press, 2012.
- [5] Gu Liping. Insight of Rural Online Shopping Consumption in China: Features, Short Board and Optimization Strategy [J]. Reform and Strategy, 2016, 32 (12): 115-117.
- [6] Lei Bing. The Relationship between Rural E-commerce Development and Local Economy -- Based on the Data of 1,870

Counties in China [J/OL]. Contemporary Economy & Management, 2018 (02): 41-47.

- [7] Yang Guang. ZhongnongOnline Shopping Is Recognized as the National Demonstration Base for Agricultural and Rural Informatization [J]. Pesticide Market Information, 2018(01):15.
- [8] Chen Qingmei. Current Situation and Countermeasures of E-commerce Rural Distribution [J]. Southern Agricultural Machinery, 2014, 48 (24): 120.