Academic Journal of Business & Management
ISSN 2616-5902 Vol. 2, Issue 5: 114-119, DOI: 10.25236/AJBM.2020.020512

# Demand Forecast of New Retail Target Products based on Multiple Linear Regression

**Beifen Zhou[1], Jinbiao Xia[1], Xinru Cheng[1], Binbin Xu[1], Yiliang Qiao[2], Xiang Wang[3] and Lan Guo[4]**

*[1] School of Information Science and Engineering, Guilin University of Technology, Guilin 541004, China*
*[2] School of Mathematics and Statistics, Qilu University of Technology, Ji'nan 250353, Shandong, China*
*[3] School of Electrical Engineering and Automation, Heilongjiang University of Science and Technology, Harbin 150027, Heilongjiang, China*
*[4] Xi'an College of Economics and Humanities, Shanghai International Studies University, Shanghai 200083, China*

***ABSTRACT.*** *Firstly, this paper obtains the SKC from July 1, 2018 to October 1, 2018 and ranks the top 50 in cumulative sales through data preprocessing. By analyzing the four holidays in 2018 and other related factors, five factors are comprehensively selected, including the actual average cost price, label price, product sales characteristics, inventory information, holiday discount, and obtains the quantitative value of each factor;The factors affecting the sales volume are calculated by the linear regression model, and then the maximum influencing factors are obtained through the linear regression analysis,The inventory information has the lowest impact on sales volume. Secondly, through data preprocessing, we get the top ten categories with the historical sales time from June 1, 2019 to October 1, 2019, and find the monthly sales volume of these sub categories from January to September, 2019, and use this month's sales volume as the original forecast data; then analyze the original data,The results show that the sales volume has a cyclical trend of rising and falling with the change of time series, so a time series forecasting model with quadratic exponential smoothing is established. Finally, the original data is brought into the model to predict the sales volume of each month in the next three months, and the predicted results of each month are obtained, and the results are brought into the MAPE formula to calculate the MAPE of the predicted value of each month in the next three months*

***KEYWORDS:*** *Data preprocessing, Multiple linear regression model, Quadratic exponential smoothing time series forecasting model*

ISSN 2616-5902 Vol. 2, Issue 5: 114-119, DOI: 10.25236/AJBM.2020.020512

## 1. Introduction

After the reform and opening up, the development of China's retail industry has gone through five stages. Until 2016, Ma Yun put forward the concept of "new retail". Since then, China's retail industry has entered a new historical development stage [1]. Each new development will face various problems, so it is necessary to predict according to the actual situation [2]. This paper establishes an appropriate mathematical model to analyze and solve the problem [3]

## 2. The establishment and solution of the first problem model

### *2.1 Establishment of multiple linear regression model*

In order to fully consider the influence of various related factors on the sales volume of the target SKC during the four holidays of national day, double 11, double 12 and new year's day in 2018, we comprehensively selected five factors: the actual average cost price, the label price, the product sales characteristics, the inventory information and the holiday discount,We established a multiple linear regression model

$$Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \eta \quad (1)$$

Where: Y represents the cumulative sales volume of each SKC product, $\beta$ Represents the coefficient, and η represents the constant term, $x_1 - x_5$ In order to represent the actual average cost price, label price, product sales characteristics, inventory information, holiday discount five factors

### *2.2 Model solving*

We can get the data of the average price, the price of the product, the price of the product, the price of the product, the price of the product, the price of the product, the price of the product, the price of the product and the price of the product

Through SPSS multiple linear regression, the R 2 is 9.337, the goodness of fit is very good, so the result is reliable, and the coefficient table is obtained. Through the analysis of the table, we can find that there is no influence factor on the label price, because it has a very weak impact on the sales volume of the target SKC, which can be ignored, so the influencing factor of tag price is excluded

Through the β coefficient of influencing factors, we can solve the multiple linear regression model

$$Y = -6.099x_1 + 64.684x_3 - 0.144x_4 + 15621.087x_5 - 13939.157 \quad (2)$$

According to the coefficient of each influencing factor in the regression model, the negative factors are the average price of actual cost and inventory information;

ISSN 2616-5902 Vol. 2, Issue 5: 114-119, DOI: 10.25236/AJBM.2020.020512

the positive factors are holiday discount and product sales characteristics. By observing the standardized β coefficient, we can find that the most influential factor on sales volume is the actual cost average price, and the lowest influence factor is inventory information

## 3. The establishment and solution of the second problem model

### *3.1 Establishment of time series prediction model based on quadratic exponential smoothing*

In order to more accurately predict the sales volume of each month of the target category in a given region in the three months after October 1, 2019, we select the sales volume of the target sub category in the nine months from January 2019 to September 2019 as the historical analysis data, so we divide it into the following steps to solve the problem:

Step 1: we select the sales volume of SKC in September 2019 based on the date column according to the sales flow table

Step 2: using the data perspective function based on Excel, taking the SKC column as the row of the data perspective and the sales volume s as the value of the data perspective, we can get all the sales volume of each SKC product in the month

Step 3: Based on the product information table, using the data matching method of Excel (vlookup), match all the sales volume of each SKC product in the previous step to the corresponding SKC product in the product information table, then each SKC product in the product information table will get all the sales volume of the month

Step 4: Based on the product information table, code the subclass (tiny)_class_The sales volume of this kind of product can be used as the data perspective value of the small month

Step 5: use the data matching method of Excel (vlookup) for the top ten target categories, and match all the sales volume of each sub category obtained in the previous step to the top ten target sub categories in the month, then all the sales volume of the top ten target sub categories in the month can be obtained

In order to visually display the historical sales volume of the target sub category from January 2019 to September 2019, we can visualize the historical sales volume of the target sub category from January 2019 to September 2019. It can be concluded that most of the target categories show an alternating pattern of rising and falling over time,In order to better predict the sales volume of each month in the next three months, we establish a quadratic exponential smoothing time series forecasting model

The second time series smoothing model is used

ISSN 2616-5902 Vol. 2, Issue 5: 114-119, DOI: 10.25236/AJBM.2020.020512

Let the time series be $y_1, y_2, \cdots, y_t$, $\alpha$ Is the weighting coefficient, $0 < \alpha < 1$ The first exponential smoothing formula is as follows:

$$S_t^{(1)} = \alpha y_t + (1-\alpha) S_{t-1}^{(1)} = S_{t-1}^{(1)} + \alpha(y_t - S_{t-1}^{(1)}) \tag{3}$$

Among them: $y_t$ For observation sequence, $S_t^{(1)}$ T is a smooth period

The recurrence formula of moving average is as follows:

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - y_{t-N}}{N} \tag{4}$$

Among them: $M_t^{(1)}$ It is a smooth exponential moving term in t period

with $M_{t-1}^{(1)}$ As $y_{t-N}$ The best estimates are as follows:

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - M_{t-1}^{(1)}}{N} = \frac{y_t}{N} + (1 - \frac{1}{N}) M_{t-1}^{(1)} \tag{5}$$

order $\alpha = \frac{1}{N}$, in order to $S_t$ replace $M_t^{(1)}$ The formula is as follows:

$$S_t^{(1)} = \alpha y_t + (1-\alpha) S_{t-1}^{(1)} \tag{6}$$

In order to further understand the essence of exponential smoothing, formula (3) is expanded in turn

$$S_t^{(1)} = \alpha y_t + (1-\alpha)[\alpha y_{t-1} + (1-\alpha) S_{t-2}^{(1)}] = \alpha \sum_{i=0}^{\infty} (1-\alpha)^j y_{t-j} \tag{7}$$

Among them: $S_t^{(1)}$ It is the weighted average of all historical data $\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, \cdots;$ Obviously there are:

$$\sum_{i=0}^{\infty} (1-\alpha)^j = \frac{\alpha}{1-(1-\alpha)} = 1 \tag{8}$$

The first smoothing index model is as follows:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha) \hat{y}_t \tag{9}$$

Among them: this formula means that the exponential smoothing value of period T is taken as the prediction value of 1 + T period

Although the primary exponential smoothing method overcomes the shortcomings of the moving average method, when the time series changes in a straight-line trend, there is still a significant lag deviation. Therefore, it must be corrected. The modified method is the same as the trend moving average method, i.e., the second exponential smoothing, this is the linear trend smoothing method

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)} \tag{10}$$

ISSN 2616-5902 Vol. 2, Issue 5: 114-119, DOI: 10.25236/AJBM.2020.020512

Among them: $S_t^{(2)}$ Is the quadratic exponential smoothing value, when the time series $\{y_t\}$ When there is a linear trend from a certain period, it can be predicted by the linear trend model similar to the trend moving average method

$$\hat{y}_{t+T} = a_t + b_t T, T = 1, 2, \cdots \tag{11}$$

$$s.t. \begin{cases} a_t = 2S_t^{(1)} - S_t^{(2)} \\ b_t = \dfrac{\alpha}{1-\alpha}(S_t^{(1)} - S_t^{(2)}) \end{cases} \tag{12}$$

Where t is the number of forecast periods; $a_t$ Is intercept; $b_t$ Is the slope, which is also called smoothing coefficient

### *3.2 Model solving*

In order to better observe the relationship between the forecast results and the original data, we plot the predicted sales volume values of the target sub categories in the three months after October 1, 2019 with MATLAB, and get the visual results. In order to obtain the MAPE of the monthly forecast value of each sub category, we first use the solving steps of this question, Obtain the actual monthly sales volume of the target category from October to December, 2019 as shown in the attached data

The target sales volume of each small month can be calculated by using MAPE's target value of 1 month after 2019

$$MAPE = \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{n * y_i} = \sum_{i=1}^{n} \frac{1}{n} * \frac{|y_i - \hat{y}_i|}{y_i} \tag{13}$$

Where: because the number of months predicted is 10, n is replaced by 10

We first calculate the MAPE value of each target subclass in each month with the formula, and then sum the MAPE values of each target subclass in each month to obtain the map of the predicted value of each month, as shown in Table 1

*Table 1 Monthly forecast MAPE*

| month | MAPE |
|---|---|
| 2019.10 | 0.068292 |
| 2019.11 | 0.130335 |
| 2019.12 | 0.096524 |

ISSN 2616-5902 Vol. 2, Issue 5: 114-119, DOI: 10.25236/AJBM.2020.020512

**References**

[1] Yang Yanxiao, Chen Yanning, Kang Suoqian, Li Rui, Yu Xinle. Analysis of tourism income sources and influencing factors based on multiple linear regression model [J]. Rural economy and technology, 2020, 31 (07): 118-120

[2] Xiao Yunmei, Liu Qiong, Qi Shuang, Liu Yi. Prediction of cold chain demand of typical agricultural products in Changsha Zhuzhou Xiangtan region based on quadratic exponential smoothing method [J]. Rural economy and science and technology, 2019, 30 (11): 158-160

[3] Xiao Yang. Coal price prediction model based on co integration theory and artificial neural network [D]. Xi'an University of science and technology, 2016