

Vegetable sales forecasting based on nonlinear programming model

Guanzong Mo^{1,*}, Youheng Guo²

¹School of Automation, Guangdong University of Technology, Guangdong, China, 510000

²Faculty of Science, Guangdong University of Petrochemical Technology, Maoming, China, 52500

*Corresponding author: 1705773676@qq.com

Abstract. The shelf life of vegetable products is relatively short, and the product deteriorates with the increase of sales with time, and most varieties are not sold on the same day. Today's domestic offline retail are under greater pressure, losses are mainly affected by changes in the fair value of financial assets, the Coronavirus and other factors. Reliable market demand analysis is especially important for replenishment and pricing decisions. In order to improve the revenue of supermarkets, this paper uses the historical sales volume and price of vegetables in supermarkets, uses the systematic clustering method to analyze the relationship between the total sales volume of each vegetable category and the cost-plus pricing with ARIMA prediction model Markov chain deepening nonlinear programming model, and gives the total daily replenishment and pricing strategy to maximize the profit of supermarket. The results of the study predicted the sales of various vegetables, and concluded that people tend to buy edible fungi and aquatic root products.

Keywords: vegetable sales, nonlinear programming model, supermarket revenue.

1. Introduction

Since reform and opening, China's vegetable industry has made great progress, playing an irreplaceable role in ensuring market supply, increasing farmer's income, and expanding trade advantages[1]. In 2019, the Institute of Agricultural Information of the Chinese Academy of Agricultural Sciences concluded through research that the supply and demand balance of the vegetable industry throughout the year is more than enough, the market operations is mainly stable, the trade surplus advantage continues to be maintained, and the price fluctuation is in line with the perennial seasonal law[2]. Nowadays, the rapid online migration of consumer consumption habits, the diversification of channels brought about by the development of e-commerce and the impact of epidemic and other factors have brought impact to stores[3], based on the historical sales volume and sales price data of sampling supermarkets, systematic cluster analysis of various vegetable sales data to grasp the degree of people's demand for various vegetables, based on non-linear programming design, supplemented by Markov chain to strengthen the forecast accuracy, give the total daily replenishment of vegetable sales and pricing strategy.

2. Method

First, data preprocessing was carried out, and 5 vegetable items with no sales records were eliminated; this paper uses the box plot, using the mean, maximum, variance, kurtosis coefficient and skewness coefficient and other indicators to analyze the sales distribution law of vegetables, because the original data has no specific data performance, can only use the grade to describe a certain phenomenon, to analyze the correlation between the phenomena, can only use the rank correlation coefficient, so the Spearman correlation coefficient is used to analyze each vegetable category in pairs[4], the sample data is too large, and the systematic clustering algorithm is used to classify the sales data of 247 vegetable items. The clustering results divide the single products into four categories, and the clustering results divide the single products into four categories, and the supermarket commodity information is preliminarily obtained. According to the obtained

information, the total sales volume of the category was used as the dependent variable and the cost-plus pricing was used as the independent variable, and the polynomial fitting was performed for 6 vegetable categories [5]. Using the seasonal ARIMA forecast model [6], the sales data of various categories from July 1 to 7 were predicted, and the basic forecast data framework of various vegetables was preliminarily obtained. Establish a nonlinear programming model, and first determine the decision variables to record the average sales unit price of each category on the current day. Get the objective function:

$$Z_j = \sum_{j=1}^6 (P_j - C_j) \times S_j \quad (1)$$

Thereinto, Z Represents total revenue. S_j Indicates the total replenishment of the category, C_j indicates the average cost of the category after considering the loss caused by vegetable loss, and its specific calculation formula is as follows:

$$C_j = (1 + \beta) \times C_j^* \quad (2)$$

In order to reduce the complexity of the model, so as to prevent the problem of infinite commodity pricing caused by overfitting and improve the generalization ability of the model, this paper adds a penalty term after the objective function to constrain. When the planned pricing is too high, the total replenishment is reduced by the constraint of penalty items; When the planned pricing is too low, the total replenishment is increased by the constraint of penalty items [7]. The expression of the penalty item is as follows:

$$\frac{P_j}{1 + \left(\frac{P_j - \mu(P_j)}{\sigma(P_j)} \right)^2} \quad (3)$$

The objective function after finally increasing the penalty term is:

$$Z_j = \sum_{j=1}^6 \max [P_j - (1 + \beta) \times C_j^*] \times S_j + \frac{P_j}{1 + \left(\frac{P_j - \mu(P_j)}{\sigma(P_j)} \right)^2}$$

$$\left\{ \begin{array}{l} S_{j\min} \leq S_j \leq S_{j\max} \\ 0 < \beta \leq 29.25\% \\ P_j^* = \left[\frac{\sum_{j=1}^6 (1 + \beta) C_j^*}{S_j} \right] \times \left[1 + \frac{(P_j - C_j^*)}{C_j^*} \right] \\ S_1 = f_1(P_1^*) = -1.062P_1^{*2} + 19.74P_1^* - 18.33 \\ S_2 = f_2(P_2^*) = 0.829P_2^{*2} - 43.44P_2^* + 511.8 \\ S_3 = f_3(P_3^*) = 0.1457P_3^{*2} - 9.984P_3^* + 163.2 \\ S_4 = f_4(P_4^*) = -1.59P_4^* + 42.29 \\ S_5 = f_5(P_5^*) = 0.4314P_5^{*2} - 18.77P_5^* + 212.2 \\ S_6 = f_6(P_6^*) = -3.691P_6^* + 101 \end{array} \right. \quad (4)$$

Solve the planning model to obtain the pricing decision, and then based on the historical sales data of each item, use the Markov chain to iterate, predict the profit of each product and rank, on the basis of the preliminary established planning model, in order to fit the actual situation, add the minimum display quantity of not less than 2.5 kg constraints, solve the planning model:

$$Z_i = \sum_{i=1}^{33} \max[P_i - (1 + \beta) \times C_i^*] \times S_i + \frac{P_i}{1 + \left(\frac{P_i - \mu(P_i)}{\sigma(P_i)}\right)^2} \tag{5}$$

$$\begin{cases} S_i \geq 2.5 \\ 0 < \beta \leq 29.25\% \end{cases}$$

Get pricing decisions.

3. Outcome

The clustering effect is shown in the following table 1:

Table 1 Descriptive statistics on sales by vegetable category

Category	Minimum	Maximum	Average value	Standard deviation	Skewness	Kurtosis
Cauliflower	0.615	5825.591	1894.610	2489.337	1.291	0.591
Flowers and leaves	0.419	27555.94	2450.493	5083.968	3.232	11.767
Peppers	1.215	28181.74	2164.703	4993.251	3.976	18.113
Solanula	4.981	11929.10	2799.499	4139.925	1.634	1.751
edible fungi	0.943	28181.74	1455.706	4017.859	5.106	30.052
Aquatic rhizomes	0.415	6059.000	1062.717	1863.646	2.287	4.352

The standard deviation of the six categories is very large, indicating that there is a significant difference between the sales volume of vegetable products in each category, showing the huge difference in the market demand of different vegetables. The skewness coefficient of the sales volume of single products included in all categories is greater than 0, indicating that the data falling to the right of the mean is too large. The kurtosis coefficient of the sales volume of single products included in flowers and leaves, peppers and edible fungi was significantly 0, indicating that the sales distribution of each single product was significantly steeper than that of the normal distribution. The kurtosis coefficient of cauliflower was close to 0, indicating that the sales distribution of each single product was relatively flat and close to normal distribution. In Figure 1, the sales pie of six kinds of vegetables shows the distribution of varieties.

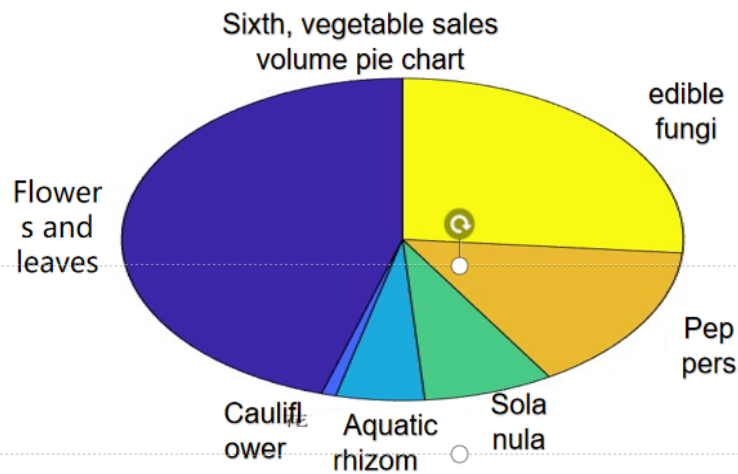


Figure 1. Six vegetable sales volume pie chart

The sales volume of the largest vegetation types is used to describe the standard distribution of the data. Figure 2 clearly shows the information of outliers, which is convenient to understand whether the data is symmetrical and how the data is grouped.

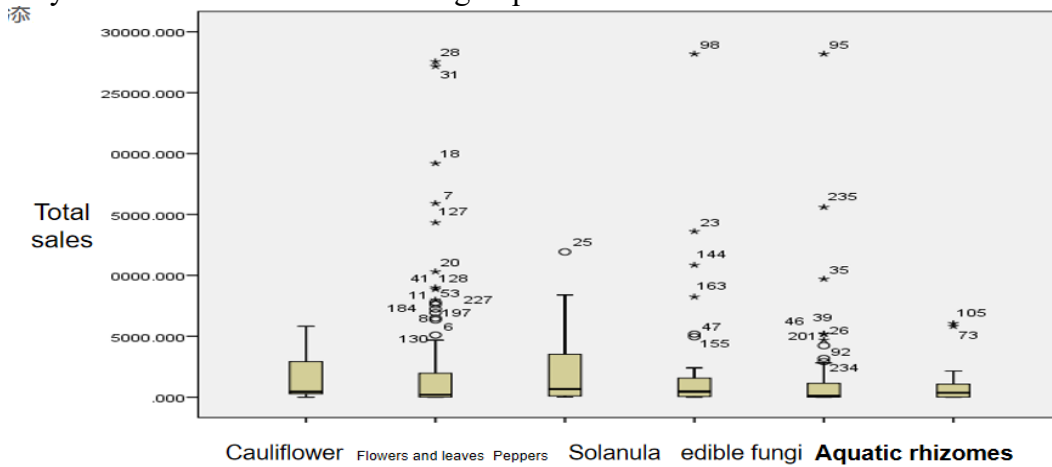


Figure 2. Sales volume of various categories of vegetables

MATLAB was used to solve the Spearman correlation coefficient of 6 categories, and Spielman correlation coefficient matrix plot for each vegetable category are shown in Figure.3. Analysis of the correlation coefficient matrix showed that there was a strong correlation between flowers and peppers, aquatic rhizomes and edible fungi, while aquatic rhizomes and nightshades had a weak correlation. There is a correlation in the sales of vegetables such as broccoli.

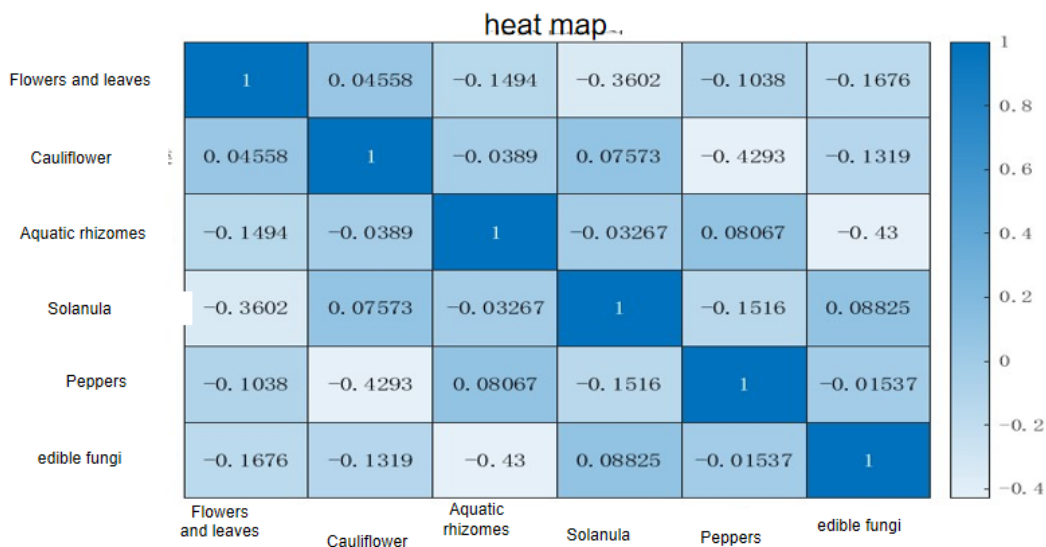


Figure 3. Spielman correlation coefficient matrix plot for each vegetable category

Table 2 Breakdown results of sales volume by vegetable unit

Category	Name
The first category	Yunnan lettuce, pure lotus root, Wuhu green pepper (1), Xishan shiitake mushroom (1), broccoli
The second category	Screw pepper, Yunnan oil wheat, purple eggplant (2), white jade mushroom (bag)
The third category	Red pepper (1), small green vegetables (1), water chestnut、、 enoki mushroom (parts), live silver fungus
The fourth category	Yellow cauliflower, spinach (parts), Shanghai green 、、 millet pepper (parts), red lotus root belt

Each vegetable item is divided into four categories, each category contains items in the table below, due to space, only some items are shown in Table 2. Among them, the sales volume of vegetable products in the first and second categories is generally high, which belongs to high-sales vegetable varieties and has high market demand. There is a correlation in the sales of vegetables such as Yunnan lettuce, pure lotus root, Wuhuqing (1), Xishan shiitake mushroom (1), and broccoli. The sales volume of the fourth type of vegetables is generally low, and the market demand is average.

Based on the ARIMA model. Through the analysis of the time series, the law of parameter change with time cycle is obtained, and the future parameter data is predicted based on this change law.

The sales forecast for July 1-7 is shown in the table 3 below:

Table 3 Sales forecast for 7.1-7.7 by vegetable category (unit: kg)

Time	Cauliflower	Flowers and leaves	Peppers	Solanula	edible fungi	Aquatic rhizomes
7.1	21.36	163.57	81.48	34.04	38.89	17.13
7.2	18.70	164.93	80.67	31.83	38.20	17.64
7.3	17.61	166.91	79.87	19.78	37.51	11.73
7.4	17.11	160.41	79.06	16.95	36.83	13.76
7.5	16.85	162.20	78.25	22.09	36.14	17.57
7.6	16.67	167.98	77.45	25.57	35.46	16.15
7.7	16.53	172.07	76.64	31.29	34.77	17.66

After analyzing the wholesale price of each vegetable unit in different time periods, it can be seen that the wholesale price of vegetables also shows cyclical fluctuations with time, so the seasonal

ARIMA prediction model can also be used to predict the average wholesale price of each vegetable unit. The final prediction results are shown in the following table 4.

Table 4 Forecast results of average prices of 7.1-7.7 for each vegetable category (unit: yuan/kg)

Time	Cauliflower	Flowers and leaves	Peppers	Solanula	edible fungi	Aquatic rhizomes
7.1	7.83	2.96	4.50	4.66	7.94	12.55
7.2	7.76	2.93	4.50	4.49	7.60	12.57
7.3	7.69	2.91	4.50	4.47	8.02	12.59
7.4	7.62	2.88	4.50	4.39	7.76	12.62
7.5	7.55	2.86	4.50	4.37	8.14	12.64
7.6	7.48	2.83	4.50	4.31	7.91	12.66
7.7	7.41	2.81	4.50	4.29	8.25	12.69

The above model was solved by using MATLAB, and the average pricing values of each vegetable category from July 1 to July 7 were obtained in the following table 5.

Table 5 Average pricing value of 7.1-7.7 for each vegetable category (unit: yuan/kg)

Time	Cauliflower	Flowers and leaves	Peppers	Solanula	edible fungi	Aquatic rhizomes
7.1	11.65	4.86	7.40	7.33	12.58	18.72
7.2	11.74	4.82	7.21	7.06	12.10	19.16
7.3	11.70	4.83	7.20	7.05	12.82	19.36
7.4	11.51	4.74	7.21	6.91	12.41	19.21
7.5	11.35	4.69	7.23	6.87	12.99	19.12
7.6	11.18	4.64	7.29	6.77	12.59	19.01
7.7	11.09	4.61	7.27	6.74	13.14	19.08

Then, according to the Markov sequence prediction, MATLAB was used to solve the above model, and the pricing strategies of 33 vegetable items on July 1 were obtained, and the results were shown in the following table 6.

Table 6 Pricing decision for selling vegetables on July 1

Name	Total sales price(Unit: RMB)	Total sales(Unit: kg)
Honghu lotus ribbon	74.65	5.51
Seafood mushrooms (bags)(4)	190.90	2.51
Hongshan vegetable treasure tote bag	99.51	4.84
Black boletus	104.30	3.54
Suizhou bubble blue	66.76	6.04
Shanghai green pepper	57.10	42.12
Millet pepper	174.69	5.42
Chicken fir nigna	121.98	3.14
Southern oil wheat	44.98	4.12
Screw pepper (servings)	29.11	11.4
Yellow cauliflower	110.75	5.67
Colorful pepper (2)	54.20	7.24
Artemisia annua	32.66	4.56
Millet pepper (servings)	2.08	7.24
Colorful pepper (1)	52.52	4.56
Sichuan red toon	79.17	2.89
Yunnan oil wheat	32.77	10.15
Gaogua(1)	34.98	9.54
spinach	33.15	3.45
Screw pepper	6.57	10.54
Cordyceps flowers	98.50	3.65
Cai Dian Quinoa Artemisia	37.25	6.45
Chicken fir fungus	110.34	17.54
Houttuynia	29.73	7.32
Branched green stems scattered flowers	41.77	9.54
Xixia Shiitake Mushroom(2)	36.68	8.41
amaranth	23.86	6.45
Silver fungus (flower)	17.07	24.51
Sweet potato tip	31.12	17.35
Sweet cabbage moss (bag)	29.48	5.12
Lotus Root(1)	13.84	6.31

Honghu lotus root belt 5.51 kg, seafood mushroom (bag) (4) 2.51 kg, Hongshan cabbage treasure tote bag 4.84 kg, black boletus 3.54 kg, Suizhou bubble green 6.04 kg, Shanghai green pepper 42.12 kg, millet pepper 5.42 kg, black skin fir fungus 3.14 kg, Nanyou wheat 4.12 kg, screw pepper (part) 11.4 kg, yellow cauliflower 5.67 kg, colorful pepper (2) 7.24 kg, foreign artemisia 4.56 kg, millet pepper (part) (part) 7.24 kg, colorful pepper (1) 4.56 kg, Sichuan red toon 2.89 kg, Yunnan oil wheat 10.15 kg, high melon (1) 9.54 kg, spinach 3.45 kg, screw pepper 10.54 kg, cordyceps flower 3.65 kg, Cai Dian quinoa 6.45 kg, chicken fir fungus 17.54 kg, houttuynia cordata 7.32 kg, Zhijiang green stem scattered flowers 9.54 kg, Xixia shiitake mushroom (2) 8.41 kg, amaranth 6.45 kg, silver fungus (flower) Purchase pricing strategy of 24.51 kg, sweet potato tip 17.35 kg, sweet cabbage moss (bag) 5.12 kg, net lotus root (1) 6.31 kg.

4. Conclusion

Through the in-depth analysis of the processing and modeling of the historical sales data of supermarkets, it is found that among the six major vegetable categories, people tend to buy edible fungi and aquatic roots, and the pricing and purchase sales strategies of supermarkets for various

commodities fluctuate periodically by seasons. Trapped by the inevitable lack of market sample collection data, the study has limitations in considering the factors of vegetable purchase price decision-making. This study reasonably plans the purchase and sales volume of supermarkets, reduces the loss of vegetable commodities, and provides a reference for purchasing decisions to maximize the profits of supermarkets in the market. The study can help improve the profit of vegetable management, improve the comprehensive operating efficiency of the market, and increase market vitality[8].

Reference

- [1] Analysis of the current situation and development prospects of China's vegetable industry in 2020 The overall price level may decrease steadily throughout the year_Research Report-Prospective Industry Research Institute (qianzhan.com)
- [2] Atul C,Ravinder S,Samriti, et al. Profitability and economic efficiency of input use in vegetable cultivation in Himachal Pradesh[J]. Indian Journal of Ecology,2021,48(2).
- [3] Surinder S,B.S. H. Sustainability of Vegetable Farming Self-Help Groups in Himachal Pradesh[J]. International Journal Of Social Sciences,2021,10(2).
- [4] Zhang Wen Yao. The degree correlation of the network is measured by Spearman's coefficient[D].University of Science and Technology of China,2016.)
- [5] Han Junhua, Gan Shengdao. Introduction to cost-plus pricing method[J].Finance and Accounting Monthly,2012.22.034.)
- [6] MA Jing, WANG Mei, CHEN Xiaowei, et al. Application of ARIMA seasonal model in predicting the incidence of hepatitis C in Lanzhou City[J].China Health Statistics,2022,39(01):98-100+105.)
- [7] HAO Wenbin. Adaptive LASSO with improved penalty term and its application in time series model[D].Anhui University of Finance and Economics,2023.000223.)
- [8] Shah P,Ansari A M. A Study of Marketing and Production Constraints Faced by Vegetable Growers[J]. Asian Journal of Agricultural Extension, Economics & Sociology,2020.