Financial Engineering and Risk Management (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/ferm.2024.070311
ISSN 2523-2576 Vol. 7 Num. 3

Research on Automatic Pricing and Replenishment Decision of Vegetable Products Based on Optimization Models

Xuanyi Jin, Xinhong Liu*, Yanqi Zhang, Ruisheng Zhang, Nana Zhao

Beijing Institute of Petrochemical Technology, Beijing, 102617, China
**Corresponding author: liuxinhong@bipt.edu.cn*

Keywords: Optimization model, genetic algorithm, time series, exponential smoothing model

Abstract: Due to the short shelf life of vegetable products, supermarkets usually make daily replenishment plans based on historical sales data and demand for various categories of vegetables. Therefore, automatic pricing and optimization of replenishment strategies for vegetable products have become particularly important. The proportion and seasonal trend of each variety have been analyzed based on the total sales of each vegetable category in this article. We established a time series model for the daily average prices of various vegetables, used exponential smoothing to predict wholesale prices for next week, and finally listed the objective function and related constraints for supermarket returns. We optimized the model solution using genetic algorithm. In addition, weather data, customer purchasing habits and preferences data, supply chain data, and other data can be supplemented to help supermarkets better formulate replenishment plans and adjust pricing strategies, achieving maximum economic benefits.

1. Introduction

The automatic pricing and replenishment decisions of vegetable products play an important role in modern retail industry. With the increasing demand for healthy fresh food from consumers, the optimization of supply chain management and sales strategies for vegetable products, as one of the main healthy ingredients, has become particularly important.

In response to the short shelf life and variable appearance of vegetable products, supermarkets usually replenish them daily based on historical sales and demand [1-2]. The pricing of vegetables generally adopts the "cost plus pricing" method [3-4], which adds a certain amount of profit to the cost. Supermarkets usually offer discounts on products with transportation damage and poor phase change. Therefore, for supermarkets, accurate market demand analysis is particularly important for replenishment and pricing decisions. From the demand side perspective, the sales volume of vegetable products often has a certain correlation with time. Supermarket stores can predict demand based on historical sales data and market trends, thereby arranging reasonable replenishment quantities and times, and avoiding situations of inventory surplus or shortage. From the supply side, the variety of vegetables available is relatively abundant from April to October. Due to the limited sales space of supermarkets, a reasonable sales mix has become crucial. Supermarkets need to select suitable vegetable varieties for replenishment and sales mix adjustment based on different seasons, regions, and consumer needs.

In summary, the automatic pricing and replenishment decisions of vegetable products need to comprehensively consider factors such as sales data, demand forecasting, and supply conditions. Through effective market demand analysis, supermarkets can optimize replenishment decisions and pricing strategies, improve sales efficiency, and meet consumer needs.

2. Model establishment and solution

2023 National College Student Mathematical Modeling C Question [5], providing product information for 6 vegetable categories distributed by a supermarket (251 items, 4 indicators), sales flow details for each product (878503 items, 7 indicators) and wholesale price related data (55982 items, 3 indicators) from July 1, 2020 to June 30, 2023, and recent loss rate data for each product (251 items, 3 indicators). For the given data, we first perform preprocessing, including handling missing values, outliers, and assigning values to vegetable categories.

2.1. Analyze the distribution pattern and correlation [6-8]

We summarize data on six types of vegetables and calculate the total sales revenue for each category from July 1, 2020 to June 30, 2023. Among them, the sales of leafy vegetables and edible mushroom vegetables accounted for 39.8% and 28.7% of the sales of supermarket vegetables, respectively. This indicates that among the sales of supermarket vegetables, leafy vegetables and edible mushroom vegetables have the highest sales, while other types of vegetables have lower sales.

We analyze and compare the monthly sales of leafy vegetables and edible mushroom vegetables in 2021 and 2022, and draw a sales trend chart as shown in Figure 1.

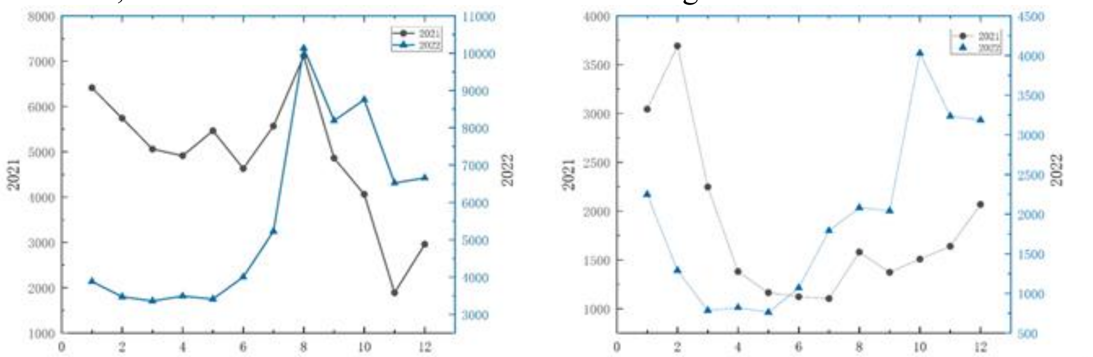

Figure 1: Sales Trends of Flower and Leaf class (Left) and Mushroom class (Right)

Figure 1 clearly shows the seasonal trend of sales volume of flowering and edible mushroom vegetables in 2021 and 2022. The highest sales volume of flowering and leaf vegetables occurred in August, which is summer. This indicates that flowering and leaf vegetables are abundant in summer, with hot weather, high temperature, sufficient sunshine, and abundant rainfall. Flower and leaf vegetables are suitable for growing in environments with high temperature, sufficient sunshine, and high humidity. Edible mushroom vegetables had the highest sales in February 2021 and October 2022, with February being winter and October being autumn. This indicates that edible mushroom vegetables are abundant in autumn and winter, with low temperatures and moderate humidity. Edible mushroom vegetables are suitable for growing in environments with lower temperatures and appropriate humidity.

We selected spinach and broccoli from the vegetable project for analysis, and based on the data from 2021, we plotted the sales unit price line chart and sales bar chart, as shown in Figure 2. Figure 2 clearly shows the changes in sales volume and unit price of two vegetables, spinach and broccoli, within a year. Spinach and broccoli are cold season vegetables with higher sales volume in autumn and the highest sales volume in October, indicating that spinach and broccoli are abundant in autumn environments with low temperatures and moderate humidity. Moreover, the sales unit price of spinach and broccoli also decreases in autumn, indicating that these two vegetables are abundant in autumn. Therefore, the distribution pattern of sales volume of spinach and broccoli shows seasonality, with high sales volume in the autumn season of high production and low sales volume in the poor harvest season. We also performed the same treatment and analysis on other vegetables.

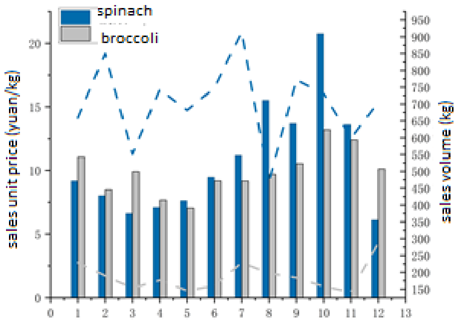

Figure 2: Sales unit price and sales volume of spinach and broccoli

The Pearson correlation coefficient is used to reflect the degree of linear correlation between two random variables. The Spearman correlation can be seen as a no parametric version of the Pearson correlation. Due to the seasonality of vegetable products, demand and price levels have a certain sequential relationship. Spearman correlation analysis is performed on all vegetable items. The partial results of the correlation coefficient with the seven colored Pepper (1) are shown in Table 1. According to Table 1, it can be seen that there is a positive and negative correlation and degree of correlation between some vegetable single products and colorful pepper (1). For example, the correlation coefficient between vegetable single product colorful pepper (2) and colorful pepper (1) is -0.313095, showing a negative correlation, indicating a certain degree of substitutability between the two. It is positively correlated with Shanghai Qingmen and Dongmenkou Xiaobaicai, indicating that they can be bundled and sold together.

Table 1: Correlation coefficient

Vegetable item name	Correlation coefficient	Vegetable item name	Correlation coefficient
Seven Colored Peppers (2)	-0.313095	Luffa tip	-0.084059
Shanghai Qingfen	0.410515	Yunnan Youmai	0.143164
Dongmenkou Cabbage	0.193854	Yunnan lettuce	0.199440

2.2. Analysis total sales volume and price

Nonlinear regression is an important statistical method used to establish nonlinear relationships between observed data. It is mainly used when the relationship between the dependent variable and one or more independent variables cannot be well described by a linear model. There is a certain correlation between the total sales volume and pricing of various categories of vegetables, and a univariate nonlinear model can be established. The general expression of the univariate nonlinear regression model is:

$$f(x_i) = a_0 x_i^0 + a_1 x_i^1 + a_2 x_i^2 + \cdots + a_n x_i^n$$

Among them, $f(x_i)$ represents the daily average pricing of the i-th vegetable category, x_i represents the daily total sales of the i-th vegetable category, and $a_0, a_1, \cdots, a_n$ represent the regression coefficients of the regression equation

The sales volume of fresh vegetables shows seasonality, with different individual vegetables being sown, matured, and harvested in different seasons. When fresh vegetables are launched in the same season, the market will experience a situation of oversupply. At this time, the sales volume and sales unit price of fresh vegetables are affected by the supply-demand relationship, and there is a non-linear relationship between sales volume and cost plus pricing. Therefore, taking the average pricing of each category of vegetables as the independent variable and the average pricing of each category as the dependent variable, a non-linear regression equation is established and solved to obtain the regression equation for each category.

The regression equation fitting parameters for various categories of vegetables are shown in Table 2.

Table 2: Model fitting parameters

Vegetable categories	MSE	RMSE	MAE	R^2
Edible fungi	0	0.003	0.002	0.992
Aquatic rhizomes	0	0.005	0.002	0.998
Solanaceae	0	0.002	0.001	0.998
Chili peppers	0	0.002	0.002	0.994
Flower leaf class	0	0.002	0.002	0.995
florescent vegetables	0	0.006	0.003	0.954

The smaller the values of MSE (mean square error), RMSE (root mean square error), and MAE (mean absolute error), the smaller the model error, and the higher the accuracy of the model. The closer R^2 is to 1, the higher the accuracy of the model. According to Table 2, the values of MSE, RMSE, and MAE are all very small, indicating that the model has a small error and high accuracy. The values of R^2 are 0.992, 0.998, 0.998, 0.994, 0.995, and 0.994, respectively, which are close to 1, indicating a high degree of regression equation fitting. In summary, nonlinear regression models have high fitting degree, high accuracy, and significant regression effect.

2.3. Time series prediction of wholesale prices

Time series prediction fully considers the randomness generated by the influence of accidental factors in the development of things. In order to eliminate the influence of random fluctuations, historical data is used for statistical analysis, and weighted average and other methods are used to appropriately process the data for trend prediction. Wholesale prices are predicted using exponential smoothing of time series data.

Exponential smoothing prediction is calculated by determining the weighting coefficient α and using the following exponential smoothing formula.

$$P_{t+1}^* = \alpha \cdot P_t + (1-\alpha) \cdot P_t^*$$

Among them, P_{t+1}^* is the $t+1$-th predicted value in the data, P_t is the t-th actual value in the data, and P_t^* is the t-th predicted value in the data.

Average the data and calculate the historical average wholesale prices from July 2020 to June

2023. We use exponential smoothing prediction method to predict wholesale prices from July 1st to July 7th, 2023. Then, the prediction error is used to determine the fitting effect of the prediction. From July 1st to July 7th, the wholesale price of aquatic roots and stems was 12.52 yuan per kilogram, while cauliflower was 7.92, 7.91, 7.90, 7.91, 7.91, 7.91, and 7.91 yuan per kilogram, respectively.

Next, the fitting degree of the model will be evaluated, as shown in Figure 3. The prediction error of each variety is controlled within 0.15, and R^2 is around 0.90, indicating that the model is relatively stable and has high accuracy. Figure 3 shows the sequence diagram of some vegetables, and it can also be seen that the predicted values are in good agreement with the actual values. The model has a high degree of fit, and the predicted data is relatively reliable.

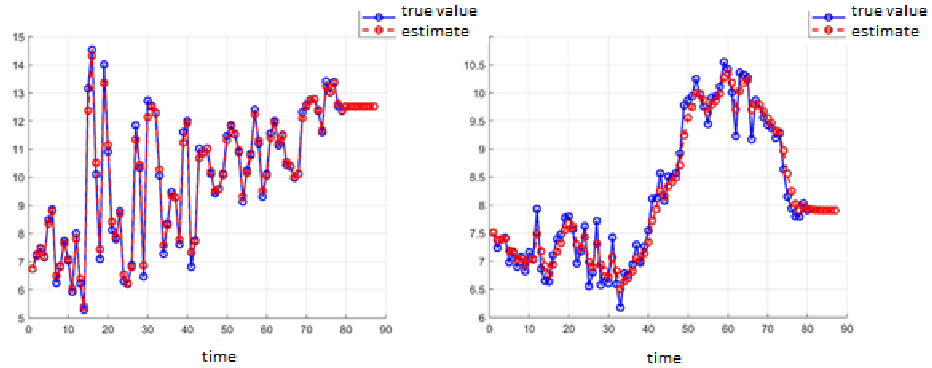

Figure 3: Prediction of aquatic rhizomes (left) and cauliflower (right)

2.4. Genetic Algorithm for Solving the Optimal Pricing Strategy and Daily Replenishment Quantity

We list the objective function based on the relationship between cost plus pricing and sales volume, provide constraints, use genetic algorithms to solve the most likely scenarios, and provide daily replenishment quantities and pricing strategies. We predict the daily replenishment volume and pricing strategy for next week (July 1st, 2023 to July 6th, 2023).

Firstly, we list the objective function. Through regression analysis, the relationship between sales volume of each category and cost markup pricing [6-9] was obtained (using the example of flower and leaf categories):

$$f_i(x_i) = 1.40 - 0.20x_i + 0.01x_i^2$$

Among them, $f_i(x_i)$ is the cost plus pricing (selling price), and x_i is the total sales volume.

The expression for the supermarket revenue of flower and leaf products is:

$$L_i = x_i \cdot f_i(x_i) - \frac{x_i}{1-h_i} \cdot P_i$$

Among them, L_i represents supermarket revenue, P_i represents wholesale price, the loss rate is h_i.

Secondly, list the following constraints:

(1) Sales volume must be greater than or equal to 0;

(2) The wholesale price should be lower than the selling price (without considering discounts) to ensure the revenue of the supermarket.

Finally, the optimization model is solved using genetic algorithm. The optimization model is:

$$\max L_i = x_i \cdot f_i(x_i) - \frac{x_i}{1-h_i} \cdot P_i$$

$$\begin{cases} x_i \geq 0 \\ f(x_i) > P_i \end{cases}$$

As this is a nonlinear programming optimization problem, if only ordinary programming methods are used, only local optimal solutions can be obtained. Therefore, a genetic algorithm is used to solve the optimization model, with 50 iterations, a population size of 30, and a mutation probability of 0.03. In order to ensure the global optimal solution is obtained, each population (each possible solution) is mutated and iterated to find the optimal solution. The specific solution results are shown in Tables 3 and Table 4.

Table 3: Strategy as of July 1, 2023

category	Supermarket revenue (yuan)	restocking volume (kg)	pricing (yuan/kg)
edible fungi	380.28	80.90	14.88
Solanaceae	105.53	26.21	9.79
florescent vegetables	272.68	71.65	12.51
Aquatic rhizomes	83.93	16.64	18.61
Flower leaf class	2280.11	512.07	8.40
Chili peppers	413.44	126.24	8.62

Table 4: Strategy as of July 2, 2023

category	Supermarket revenue (yuan)	restocking volume (kg)	pricing (yuan/kg)
edible fungi	298.89	118.33	12.29
Solanaceae	63.84	35.84	7.08
florescent vegetables	146.65	53.42	11.23
Aquatic rhizomes	79.19	24.51	16.42
Flower leaf class	843.54	287.62	6.54
Chili peppers	255.74	150.84	6.71

According to Tables 3 and Tables 4, it can be seen that the sales volume and pricing strategy for July 1 and 2, 2023 were as follows: the restocking volume of edible mushrooms on July 1 was 80.90 kg, priced at 14.88 yuan/kg, and the supermarket revenue was 380.28 yuan; The restocking volume on July 2nd was 118.33 kilograms, priced at 7.08 yuan/kilogram, and the supermarket revenue was 289.89 yuan.

2.5. Replenishment and pricing decisions for maximizing supermarket benefits based on time series and genetic algorithms [9-12]

We developed a sales strategy for July 1, 2023 based on available variety data from June 24 to 30, 2023, and solved the optimal solution under given constraints. Next, we limit the number of categories to 27-33, filter and process the data to remove the minimum sales volume of categories sold from the 24th to the 30th until there are 27 remaining individual product quantities. Then, we use exponential equations to predict and analyze the wholesale prices of individual products from the 24th to the 30th, and predict the wholesale prices on July 1st. Finally, based on section 2.5, the optimal solution is solved.

Firstly, list the objective function. Due to the need for the optimal solution for each individual product, the profit of each individual product can be summed based on the objective function in section 2.5. The objective function is as follows:

$$\max L=\sum_{i=1}^{33}\left(x_i\cdot f_i(x_i)-\frac{x_i}{1-h_i}\cdot P_i\right)$$

Among them, x_i is the sales volume of the *i*-th item, $f_i(x_i)$ is the pricing of the *i*-th item, h_i is the loss rate of the *i*-th item, and P_i is the wholesale price of the *i*-th item.

Secondly, list the following constraints:

(1) The display quantity requirement is that the sales volume must be greater than or equal to 0, and not less than 2.5 kilograms;

(2) Pricing needs to be greater than wholesale price to ensure the revenue of supermarkets;

(3) Try to meet the needs of various vegetable varieties, and the variety of individual products should be within the range of 27-33.

Based on the objective functions and constraints listed above, the following optimization models can be listed.

$$\max L=\sum_{i=1}^{33}\left(x_i\cdot f_i(x_i)-\frac{x_i}{1-h_i}\cdot P_i\right)$$

$$\begin{cases} 2.5\le x_i \\ P_i< f_i(x_i) \\ 27\le i\le 33, i\in Z \end{cases}$$

Table 5: Partial Single Product Strategy Table

Single item	Supermarket revenue (yuan)	pricing (yuan/kg)	restocking volume (kg)
Shanghai Qing	90.81	7.59	25.70
Wuhu Green Pepper (1)	66.79	7.92	42.68
Bamboo leafed vegetable	25.84	7.49	22.08
Xixia Flower Mushroom (1)	67.39	7.50	33.41
Zhijiang Qinggan Scattered Flowers	149.60	7.79	48.00

Finally, we apply genetic algorithm to solve the optimization model. We use genetic algorithm to solve each individual product separately. Among them, Shanghai Qing has 4500 iterations, a population of 100, and a mutation probability of 0.05; The iteration number of Wuhu green pepper (1) is 5000 times, the population is 100, and the probability of variation is 0.05. According to the genetic algorithm, the sales strategies for 27 individual products were calculated, and some results are shown in Table5. According to Table 5, the sales strategy of Shanghai Qing is to set the price at 7.59 yuan/kg, replenish 25.70 kg, and achieve a maximum supermarket revenue of 90.81 yuan; The sales strategy of Wuhu Green Pepper (1) is to set a price of 7.92 yuan/kg, with a restocking volume of 42.68 kg and a maximum supermarket revenue of 66.79 yuan.

3. Conclusion

The relationship between the sales unit price and sales volume of various categories of vegetables was accurately predicted through a nonlinear regression model. Predict wholesale prices of various categories of vegetables using exponential smoothing method. Solve optimization models through genetic algorithms. Genetic algorithm has good global search performance, which can explore and find the global optimal solution or near optimal solution in complex solution spaces and large data similar to this problem. Reasonable replenishment and pricing decisions can bring huge economic benefits to enterprises. Therefore, in order to better formulate replenishment and pricing decisions for vegetable products, supermarkets can consider collecting seasonal and weather data, and other related data for analysis. These data provide a lot of information about the market, consumers, and products, helping supermarkets better formulate replenishment plans and adjust pricing strategies to achieve maximum economic benefits.

Acknowledgements

The financial support from the Research on the Teaching Reform and Practice of Integrating Ideological and Political Elements into the Probability and Statistics A Course of the Key Teaching Reform Project of BIPT in 2023 (ZD202306001); Supported by the Zhiyuan Science Foundation of BIPT (No. 2024212).

References

[1] Zheng Yan. Analysis of the Implementation Status of Vegetable Marketing Strategies in Lighthouse City [D]. Shenyang Agricultural University, 2017.

[2] Mao Lisa. Research on Pricing Strategies and Production and Sales Models of Vegetable Wholesale Market from the Perspective of Supply Chain [D]. Central South University of Forestry and Technology, 2023.

[3] Han Junhua, Gan Shengdao. Introduction to Cost plus Pricing Method [J]. Finance and Accounting Monthly, 2012 (22): 74-75.

[4] Wei Zee, Chen Gang, Ding Shengchun, et al. Research on Cost plus Pricing Method for Mass Customization Products [J]. Heilongjiang Foreign Economic and Trade Journal, 2007 (02): 84-85.

[5] Question C of the 2023 National College Student Mathematical Modeling Competition [EB/OL]. The official website of the National College Student Mathematical Modeling Competition, http://www.mcm.edu.cn.

[6] Yang Haoxu. Research on the correlation between vegetable prices and sales: A case study of rapeseed [J]. Food Safety Journal, 2018 (21): 179-181.

[7] Cui Yunhao. Research on Fresh Vegetable Sales Prediction Based on CNN-PSO-LSTM Combination Model [D]. Anhui Agricultural University, 2023.

[8] Lv Wang. Research on Predicting Sales Trends of Fresh Vegetables Based on Improved SVM [D]. Anhui Agricultural University, 2022.

[9] Wang Lijuan, Liu Guifeng, Xin Liyuan, et al. Volatility and Short term Prediction of Vegetable Prices in Tianjin: Seasonal Adjustment and Prediction Based on Time Series [J]. Chinese Journal of Agriculture, 2015, 31 (09): 286-290.

[10] Tang Junjie, He Meiyi. Research on vegetable market price prediction based on genetic algorithm and Gaussian process regression [J]. Journal of Changchun University of Finance, 2017 (05): 80-87.

[11] Luo Changshou. Research on vegetable market price prediction method based on neural network and genetic algorithm [J]. Science and Technology Bulletin, 2011, 27 (06): 881-885+894.

[12] Qian Binbin, Xie Shenru, Yang Baohua. A Time Series Based Wavelet Neural Network Vegetable Price Prediction Model [J]. Journal of Luoyang Institute of Technology (Natural Science Edition), 2016, 26 (04): 63-67.