Research on Pricing and Replenishment Strategy of Supermarket Fresh Food Based on Multiple Regression and Particle Swarm Optimization

Zesen Zheng *, #, Hongbin Feng #

School of Mathematics and Statistics, Guangdong University of Technology, Guang Zhou, China, 510520

* Corresponding author: zesen1791413202@gmail.com

#These authors contributed equally.

Abstract. The characteristics of vegetable commodities such as short storage period, easy to wear and tear, and rapid demand for circulation time led to daily changes in their pricing. To develop a reasonable pricing and replenishment plan, this paper first analyzes the relationship between sales volume and pricing by using a multiple regression model, with pricing as the dependent variable and wholesale price and sales volume as the independent variables. The multivariate regression equation was derived by determining whether there is multicollinearity and heteroskedasticity among the independent variables through the variance inflation factor and White's test, and OLS + robust standard error regression was used to reduce the effect of heteroskedasticity. Secondly, replenishment and pricing strategies were developed for different categories, firstly, time series analysis was conducted to predict the cost, then particle swarm optimization was used to find out the optimal pricing, and based on the relationship between the pricing and the sales volume, the sales volume was projected, and finally the replenishment volume was inverted. The study provides practical suggestions and references for pricing and replenishment decisions in the supermarket fresh food industry.

Keywords: Multiple Regression, Particle Swarm Optimization, Fresh Pricing Strategy.

1. Introduction

With the rapid development of the economy, the sales and distribution channels of vegetables and other fresh food have undergone great changes, and at the present stage of China's fresh food circulation, the proportion of fresh food superstores is increasing. For fresh food superstores, fresh food is characterized by a short storage period, easy wear and tear, and rapid demand for circulation time, which leads to price changes that may occur within hours or even minutes. Supermarkets at the end of the supply chain should consider not only the characteristics of fresh food but also other external factors when determining the price of fresh food [1].

To develop a reasonable pricing and replenishment plan, the Pearson correlation model of vegetable category and sales distribution can be established and solved to establish an ARIMA prediction model to make a prediction of the pricing of the goods [2]; and with the development of artificial intelligence and 5G technology, there are also “online + offline” sales strategies based on 4Ds [3] and based on the 4P marketing theory and the SICAS model of the sales strategy [4]. In addition, to study the pricing strategy, bilateral quality control strategy and government subsidy strategy for profit maximization of fresh produce supply chain under different market environments and under the domination of different subjects, it can be helped by methods and research tools such as demand function theory, price theory, game theory, utility theory, optimization theory, supply chain coordination contract theory and numerical calculation [5-7].

Returning to this paper, we first use regression analysis to get the relationship between sales and cost-plus pricing. Then the time series analysis predicts the cost for the coming week, and then the particle swarm optimization is used to solve what the optimal pricing is. Once we have the pricing, we can get the specific sales volume based on the relationship equation from the regression analysis, and then from the sales volume we can inversely derive the replenishment volume. This completes the process of developing a pricing and replenishment strategy that maximizes revenue.

2. Solving for the relationship between sales volume and pricing

2.1. Data pre-processing

For the given data, since we are required to analyze the relationship between total sales and cost-plus pricing for each vegetable category, and discounted data is often due to shipping losses and deterioration in the quality of the vegetable category, which can affect the analysis of the above relationship, we have decided to exclude all the discounted data in this problem. The data for this article comes from: http://www.mcm.edu.cn/html_cn/node/c74d72127066f510a5723a94b5323a26.html

2.2. Modeling and solving

There are many factors affecting the pricing of fresh food in supermarkets, especially the impact of the wholesale price and sales volume of the day on the pricing should not be underestimated, so we consider the wholesale price and sales volume as the independent variables, and pricing as the dependent variable, and use multiple regression to explore the relationship between them. Figure 1 shows the flow chart of multiple regression.

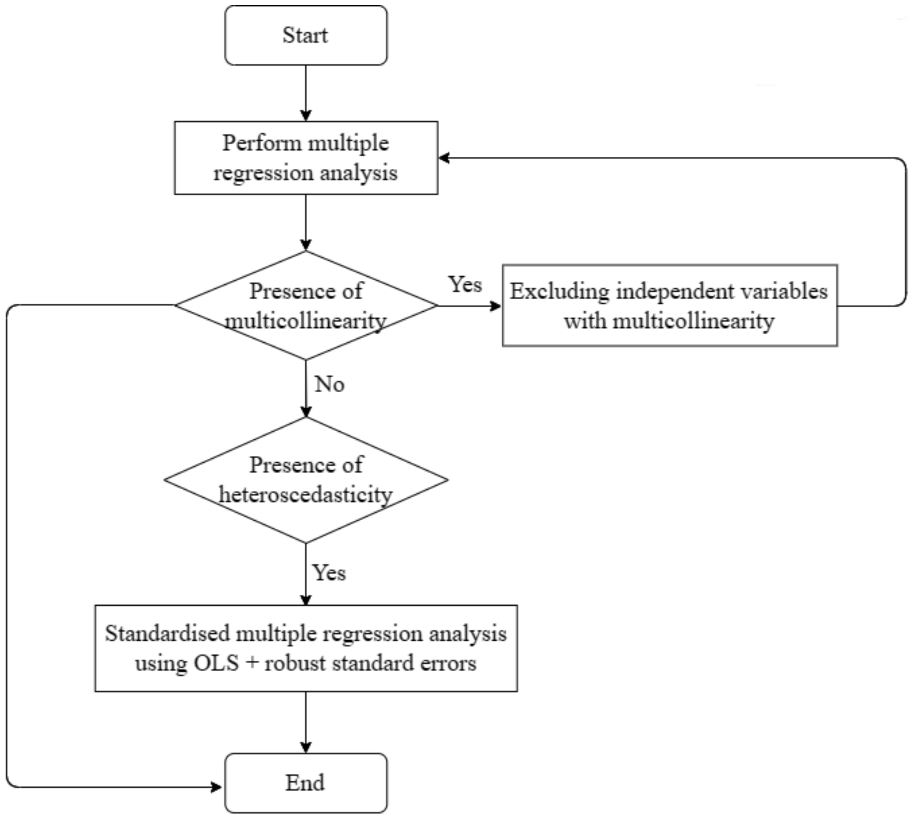

Figure 1. Flowchart of multiple regression analysis

Step1: We need to judge whether there is multicollinearity between wholesale price and sales volume among categories, mainly through Variance Inflation Factor (VIF), now assume that there are k independent variables, then the mth independent variable's Variance Inflation Factor, VIF_m, is shown in equation (1).

$$VIF_m = \frac{1}{1-R^2_{1\sim k/m}} \tag{1}$$

$R^2_{1\sim k/m}$ is the fit obtained by regressing the mth independent variable as the dependent variable on the remaining k-1 independent variables. The larger VIF_m is, the larger $R^2_{1\sim k/m}$ is, indicating that the mth variable is highly correlated with other variables. If the VIF of the multiple regression model is defined as:

$$VIF = max\{VIF_1, VIF_2, \dots, VIF_m\} \tag{2}$$

Table 1 shows the results of the multicollinearity test, which indicates that none of the categories are multicollinear.

Table 1. Results of the multicollinearity test

	Cauliflower	Philodendron	Chilli	Eggplant	Edible fungi	Aquatic rhizomes
VIF	1.01	1.08	1.03	1.04	1.02	1.11

Step2: Use White test to determine whether there is heteroskedasticity, assuming the original hypothesis of White test H0: there is no heteroskedasticity in the model. After the calculation of STATA software, the p-value of all categories is less than 0.05, so the original hypothesis is accepted and there is no heteroskedasticity.

Step3: Assuming that x_1 is the sales volume, x_2 is the wholesale price, and y is the pricing, the multiple regression results are shown in Table 2.

Table 2. Results of multivariate regression

Vegetables	Multiple regression
Cauliflower	$y = -0.0008x_1 + 1.2486x_2 + 1.9333$
Philodendron	$y = -0.0087x_1 + 1.2658x_2 + 2.5099$
Chilli	$y = 0.0045x_1 + 1.3446x_2 + 1.8391$
Eggplant	$y = 0.0053x_1 + 1.3589x_2 + 1.3295$
Edible fungi	$y = 0.0032x_1 + 1.1735x_2 + 2.0879$
Aquatic rhizomes	$y = 0.0009x_1 + 1.1668x_2 + 2.5646$

3. Replenishment and pricing strategies

Since the formula for the daily profit of a supermarket is shown in (3), to explore the '$Cost$' in the formula, i.e. the wholesale price, we first use time series analysis to predict the future cost.

$$DailyProfit = DailySales \cdot Pricing - DailyReplenishment \cdot Cost \tag{3}$$

3.1. Time series analysis of forecasting costs

Time series refers to a phenomenon of indicator values in accordance with the time order of the numerical sequence, therefore, there must be some kind of law behind the numerical changes in the time series, these laws are the time series analysis, which is also an important entry point for this paper.

Step1: In general. Time series can be decomposed into the following four kinds of laws: long-term change law (T), seasonal change law (S), cyclical change law (C), irregular change (I). In general, time series are often the result of the superposition of these patterns. Since the given data is only three years old, we decided not to consider the effect of cyclical patterns. In the following, we take the quarterly sales of 'eggplant' as an example and make a time series plot (time series plot) of it, as shown in Figure 2. It is obvious that the time series plot of eggplant has a clear seasonal variation pattern.

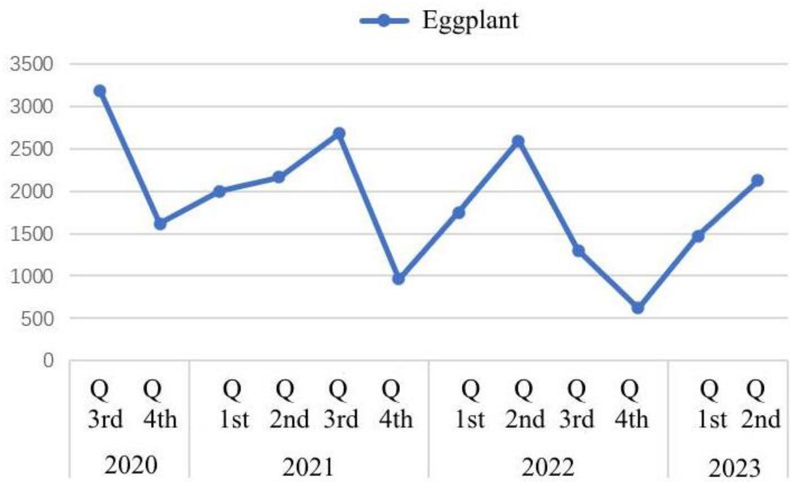

Figure 2. Time series plot of eggplant

Step2: If a time series has four regular movements, and if all four movements are independent of each other, then the superposition model can be expressed as:

$$Y = T + S + C + I \tag{4}$$

If there is an interaction between the four changes, then a product model should be used:

$$Y = T \cdot S \cdot C \cdot I \tag{5}$$

Step3: According to the different patterns presented by different time series, time series analysis requires the establishment of a suitable model to solve. Common models include: Winters' additive model (Winters' additive), Simple seasonal model (Simple seasonal), ARIMA (p, d, q) model and so on [8]. Among them, Winter's additive method is applicable to the series with linear trend and seasonal effect does not change over time; Simple seasonal model is applicable to the series with no trend and the seasonal effect remains constant over time; ARIMA (p, d, q) model is known as autoregressive differential moving average model, which is a kind of autocorrelation model that attempts to solve the problem by using the autocorrelation of the data to improve the quality of the data. The ARIMA (p, d, q) model is called the autoregressive differential moving average model, which is a model that tries to extract the hidden patterns of time series through the autocorrelation and differentiation of the data [9].

The overall flowchart of the time series analysis is shown in Figure 3.

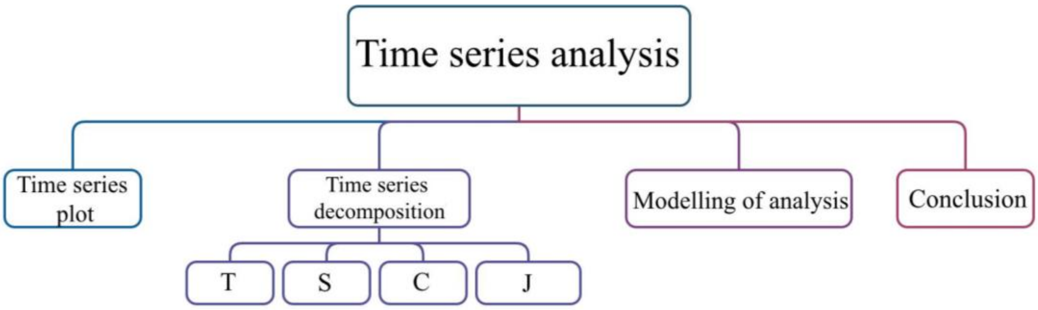

Figure 3. Flowchart of time series analysis

Step4: Since the wholesale price given in the data is by individual product, and we require analysis by category, the wholesale price of the category is averaged uniformly. According to the time series analysis, the category cost from 1 to 8 July 2023 is shown in Table 3.

Table 3. Cost forecasts

Date	Cauliflower	Philodendron	Chilli	Eggplant	Edible fungi	Aquatic rhizomes
2023/7/1	7.73	3.53	5.94	4.43	4.70	11.87
2023/7/2	7.71	3.53	5.94	4.43	4.70	11.87
2023/7/3	7.70	3.54	5.94	4.43	4.70	11.83
2023/7/4	7.70	3.55	5.94	4.43	4.70	11.80
2023/7/5	7.68	3.55	5.94	4.43	4.70	11.78
2023/7/6	7.67	3.56	5.94	4.43	4.70	11.77
2023/7/7	7.66	3.57	5.94	4.43	4.70	11.77

3.2. Pricing strategy

Equation (3) can be described in mathematical terms: assume that pricing is x and profit is z, according to the multivariate regression in Section 2.2, we know that there is a relationship between pricing and sales volume, so let's set it to $f(x)$, then assume that all the products are sold on that day, so the daily replenishment can be obtained based on the data given by the attrition rate (l) and the sales volume, and according to the cost (u) in Section 3.1, it is also clear. Therefore, equation (3) is rewritten as equation (6).

$$z = xf(x) - u\frac{f(x)}{l} \tag{6}$$

From the data, it can be obtained that the pricing x fluctuates within a specific range, to find the optimal x to maximize the profit z, we consider that the essence is the optimisation problem, therefore the particle swarm optimization of intelligent optimization algorithms is used to solve the problem.

3.3. Particle swarm optimization for finding optimal solutions

Particle swarm optimization, also known as PSO. It is a search algorithm based on group collaboration developed by simulating the foraging behavior of bird flocks [10]. Its specific flowchart and pseudo-code are shown in Figure 4 and Figure 5.

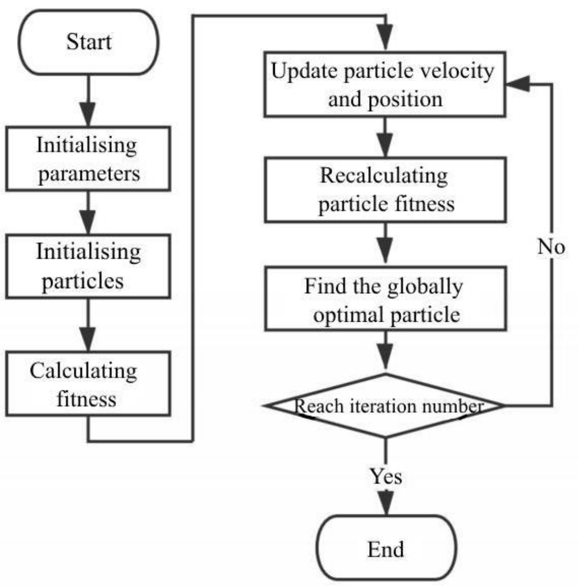

Figure 4. PSO Flowchart

Algorithm 1 PSO

```
for each partical i do
  Initialize velocity V_i and position X_i for partical i
  Evaluate partical i and set pBest_i = X_i
end for
gBest = min {pBest_i}
while not stop do
  for i = 1 to N do
    Update the velocity and position of partical i
    Evaluate partical i
    if fit(X_i)<fit(pBest_i) then
      pBest_i = X_i
    end if
    if fit(pBest_i)<fit(gBest) then
      gBest = pBest_i
    end if
  end for
end while
print gBest
```

Figure 5. PSO Pseudo-Code

The core equations are shown in equation (7) and (8).

$$v_i^d = wv_i^{d-1} + c_1r_1\left(pBest_i^d - x_i^d\right) + c_2r_2\left(gBest^d - x_i^d\right) \quad (7)$$

$$x_i^{d+1} = x_i^d + v_i^d \quad (8)$$

Where d denotes the number of iteration steps, v_i^d denotes the velocity of the i^{th} particle at the d^{th} iteration, w denotes the inertia weight, also known as the inertia coefficient, c_1 denotes the individual learning factor of the particle, c_2 denotes the social learning factor of the particle, r_1, r_2 is a random number on [0, 1], $pBest_i^d$ denotes the best position passed by the i^{th} particle up to the d^{th} iteration, and $gBest^d$ denotes the best position passed by all particles up to the d^{th} iteration. According to general experience, it is more appropriate to take 2 for the individual learning factor and the social learning factor, and it is more appropriate to take 0.9-1.2 for the inertia weight, which is generally taken as 0.9.

Equation (6) is the objective function in the PSO algorithm, and we use the particleswarm function in Matlab to solve for the maximum value in the multivariate regression equation to obtain the optimal pricing as shown in Table 4.

Table 4. Pricing strategy for each category

Date	Cauliflower	Philodendron	Chilli	Eggplant	Edible fungi	Aquatic rhizomes
2023/7/1	10.06	5.44	13.78	9.07	14.38	10.00
2023/7/2	10.06	5.44	13.78	9.07	14.38	10.00
2023/7/3	10.06	5.45	13.78	9.07	14.38	10.00
2023/7/4	10.06	5.46	13.78	9.07	14.38	10.00
2023/7/5	10.00	5.47	13.78	9.07	14.38	10.00
2023/7/6	9.99	5.47	13.78	9.07	14.38	10.00
2023/7/7	9.98	5.48	13.78	9.07	14.38	10.00

3.4. Replenishment strategy

We derive the sales volume from the pricing information in Table 4 and the multiple regression equation (Table 2), and then divide the sales volume by the attrition rate to obtain the stocking volume, or replenishment volume. It is worth noting that since the data gives the attrition rate for a single item, we decided to express the attrition rate for the category in terms of the median. The results are shown in Table 5.

Table 5. Replenishment strategy for each category

Date	Cauliflower	Philodendron	Chilli	Eggplant	Edible fungi	Aquatic rhizomes
2023/7/1	23.00	195.68	86.00	17.49	38.60	16.50
2023/7/2	22.29	195.74	75.16	18.37	38.63	18.16
2023/7/3	21.71	195.80	72.28	18.67	36.98	15.42
2023/7/4	20.04	195.87	74.32	16.61	37.10	15.31
2023/7/5	21.50	195.83	75.61	16.87	38.99	12.69
2023/7/6	22.50	196.00	73.40	15.80	44.43	12.09
2023/7/7	24.55	196.06	76.79	18.11	44.27	13.51

4. Conclusion

With the increasing diversification of consumer demands and fierce competition in the market, superstores, as an important part of the retail industry, are facing unprecedented challenges and opportunities. To maintain market competitiveness while achieving business objectives and maximizing profits, stores therefore need to strengthen the research and practice of replenishment and pricing strategies, and formulate scientific and reasonable replenishment and pricing strategies to cope with changes in the market and changes in consumer demand. In the practical application of replenishment and pricing strategies, superstores need to focus on the coordination and cooperation between the two. On the one hand, the development of replenishment strategy needs to consider the impact of the pricing strategy, to ensure that the replenishment can consider the sales of goods in different price ranges and inventory demand; on the other hand, the development of pricing strategy needs to consider the implementation of the replenishment strategy, to ensure that the pricing can be considered the cost of goods inventory and changes in market demand.

In addressing the intricate challenges of pricing and replenishment in the retail sector, this paper delves into the intricate relationship between sales volume and pricing methodologies. Employing a multiple regression model as the foundation for analysis, we employ two statistical techniques — variance inflation factor and White's test — to carefully examine whether the independent variables exhibit multicollinearity or heteroskedasticity. To mitigate the potential impact of heteroskedasticity, we adopt the robust standard error regression method alongside OLS, ultimately arriving at a refined multiple regression equation. The outcomes of this rigorous analysis are presented in Table 2.

Furthermore, to formulate tailored replenishment and pricing strategies for various product categories, we first leverage time series analysis to forecast costs accurately. Subsequently, we harness the power of the particle swarm optimization to derive the optimal pricing data, as presented in Table 4. Based on the established relationship between pricing and sales volume, encapsulated in Equation (6), we project sales volumes and subsequently deduce the required replenishment quantities. The culmination of this comprehensive analysis is presented in Table 5, offering a comprehensive view of the optimized replenishment and pricing strategies.

The research on the issue of superstore replenishment and pricing strategy is of great significance for superstore operation and management, enhancing market competitiveness, meeting consumer demand, promoting supply chain synergy and promoting industry innovation. Therefore, we should strengthen the research and exploration of superstore replenishment and pricing strategies to cope with the changing market environment and consumer demand.

References

[1] Wang F. Research on Dynamic Pricing Strategy of Supermarket Fresh Food [D]. Southeast University, 2007.

[2] Nie Yuxuan. Research on automatic pricing and replenishment strategy of fresh commodities based on ARIMA prediction optimization model - taking vegetable commodities as an example [J]. Commercial Exhibition Economy, 2024 (05): 19 - 22.

[3] Chen Dahong. Research on marketing strategy of Park supermarket based on 4Ds[D]. Lanzhou University of Technology, 2023.

[4] Teng Zichen. Research on new retail marketing strategy of Freshippo based on SICAS model [J]. Journal of Economic Research, 2021, (36): 50 - 52+60.

[5] Wen Hui. Research on profit maximization strategy of fresh agricultural products supply chain under the domination of different subjects [D]. Huazhong Agricultural University, 2021.

[6] Zhang Jingling. Research on differential pricing strategy considering consumer choice behavior [D]. Nanjing University of Information Engineering, 2022.

[7] Ding Yirang. Research on fruit synergistic pricing strategy considering reference price and competitive effect [D]. Beijing University of Chemical Technology, 2021.

[8] Lu Biao. Application of Time Series and Decision Tree Modelling in Online Hotel Sales Forecasting [D]. Nanchang University, 2023.

[9] Wu Weizhong, Deng Zhong, Chen Yudao, et al. Characterizations of time series variation of water influx in Panlong lead-zinc mine in Guangxi and ARIMA prediction [J]. Journal of Guilin University of Technology, 2023, 43 (03): 406 - 413.

[10] Liu Jianghua. The basic theory of particle swarm optimization and its improvement research [D]. Central South University, 2010.