

# A Study on Pricing and Replenishment Decision of Vegetable Commodity Based on Demand-Volume-Cost Model

Xufeng Li, Qingxin Huang \*

School of Mathematical Science, South China Normal University, Guangzhou, China, 510631

\* Corresponding Author Email: 20212232018@m.scnu.edu.cn

**Abstract.** This paper aims to develop pricing and replenishment strategies for vegetable commodities in superstores. The characteristics of perishability, short life cycles, and other factors make it challenging to determine pricing and replenishment decisions. The study proposes the use of a weighting method to define the price for each vegetable category. Additionally, sub-models are established for market demand prediction, price-volume relationship, and cost-plus pricing. The ARIMA time series forecasting model is used for market demand forecasting, while a logarithmic model, fitted non-linearly using the L-M algorithm, is employed for the price-sales model. Through the derived sub-model expressions, the optimal replenishment volume and markup rate that maximize the superstore's profit are determined. The calculated maximum profit from selling vegetable goods in the coming week is 2179.68 yuan. In this paper, a profit model is constructed by considering the relationship between demand, sales volume, and cost, which can provide suggestions for superstores to improve the economic efficiency of vegetable products.

**Keywords:** LM algorithm, Arima model, Planning model, Optimization model.

## 1. Introduction

Fresh produce refers to the farm or farmer directly sown, responsible for the care until maturity, without a series of processing of the product action, and in the normal environment cannot be stored for a long time of a class of agricultural products, usually contains fruits and vegetables and beef and lamb and fish, shrimp, and other aquatic products, etc. [1]. Vegetables, as an important category of fresh agricultural products, are an indispensable part of people's daily diet, and have an important impact on China's economic development and social stability [2]. Unlike ordinary commodities, fresh agricultural products have the characteristics of perishable, easy to deteriorate, short life cycle, and so on.

In order to better improve the economic efficiency of such commodities, scholars have done a lot of research: Yali Li proposed a strategy of reasonable investment in preservation technology to reduce the cost of spoilage [3]; Aonan Liang and Shuyun Wang constructed a two-phase profit model for pre-sale and regular sales of fresh agricultural products, so as to optimize the decision-making on pre-sale time, two-phase pricing and production scale [4]; Zhigang Lu and Cheng Chen based on consumer choice behavior to classify customers based on consumer choice behavior and investigated the optimal pricing strategy when retailers adopt a separation model for the sale of fresh farm products of both high and low quality levels[5]; Hao, Yangyang, and Wrong, Yu optimized the BP neural network forecasting model using the IPSO algorithm; thereby constructing a logistics demand forecasting model for fresh farm products with higher accuracy [6]; Yao Chun and Lv Huiyan established a generalized function of freshness that decays over time and jointly affects the demand for fresh produce with price to study and analyze the pricing strategy of a certain type of fresh produce under different wastage and discount factors [7]; HAN et al; He et al constructed a retailer profit model based on the quality grade of fresh food, and thoroughly investigated the impact of optimal grading of fresh produce on the retailer's optimal pricing [8].

The above studies have laid a good theoretical foundation for the study of pricing and replenishment decisions of vegetable commodities, but very few studies have been able to explore in great depth the relationship between demand, sales volume, and cost, as well as their impact on the pricing and replenishment decision problem of vegetable commodities. Based on previous studies,

this paper constructs a demand-volume-cost model for pricing and replenishment decisions of vegetable commodities to maximize the profit of merchants.

## 2. The basic fundamental of model

### 2.1. L-M algorithm

The L-M (Levenberg-Marquardt) algorithm is an optimization algorithm commonly used for nonlinear minimization problems. The central idea is to find the optimal parameters of a function to minimize a given loss function. This loss function usually represents the difference between the model predictions and the actual observations, and the goal is to minimize this difference by adjusting the model parameters. The L-M algorithm combines the method of the most rapid descent with the method of linearization (Taylor series), which allows us to find the optimal value relatively quickly [9].

Before introducing the LM algorithm, it is necessary to introduce the Newton method. The iterative formula for Newton's method is:

$$x_{k+1} = x_k - (J(x_k)^T J(x_k) + S(x_k))^{-1} J(x_k)^T F(x_k) \quad (1)$$

where,

$$S(x) = \sum_{i=1}^m F_i(x) \nabla^2 F_i(x) \quad (2)$$

Since the  $\nabla^2 F_i(x)$  of the Hessian matrix of the residue function  $F_i(x)$  in  $S(x)$  is more computationally intensive, if this item is ignored, the Gauss-Newton iterative formula is immediately obtained:

$$x_{k+1} = x_k - (J(x_k)^T J(x_k))^{-1} J(x_k)^T F(x_k) \quad (3)$$

However, the Gauss-Newton method needs to require the matrix  $J(x_k)^T J(x_k)$  to satisfy the full rank of the columns in the process of specific application, and this limitation constrains its application. The LM algorithm, on the other hand, obtains the search direction by solving the following optimization model:

$$d_k = \arg \min ||F(x_k) + J_k d||^2 + \lambda_k ||d||^2 \quad (4)$$

Eq. (4) in which  $\lambda_k > 0$ . From the optimality condition,  $d_k$  satisfies equation(5):

$$\nabla (||F(x_k) + J_k d||^2 + \lambda_k ||d||^2) = 0 \quad (5)$$

The result is:

$$d_k = -(J_k^T J_k + \lambda_k I)^{-1} J_k^T F(x_k) \quad (6)$$

The final iterative formulation of the L-M algorithm is obtained as in equation (7):

$$x_{k+1} = x_k - (J_k^T J_k + \lambda_k I)^{-1} J_k^T F(x_k) \quad (7)$$

It can be seen that the key to the L-M algorithm is the need to adjust the parameter. When  $\lambda_k$  is too large, the iteration speed of the algorithm decreases and it is less efficient; when  $\lambda_k$  is too small, the matrix  $J_k^T J_k + \lambda_k I$  is close to the singular, and the algorithm is prone to divergence and less stable.

### 2.2. ARIMA model

The definition of the time series model ARIMA (p, d, q) model is: if the sequence after a d-order process  $x_t$  is stationary, that is, satisfies the difference equation:

$$\Phi(B)\nabla^d X_t = \theta(B)\varepsilon_t \quad (8)$$

Where  $d$  is the order of difference. If  $\Phi(B)$  and  $\theta(B)$  are  $p$ -order and  $q$ -order polynomials respectively, then ARIMA ( $p, d, q$ ) model with  $X_t$  following the difference order  $d$ .  $p, q$ -metric stationary sequence the short-term memory of  $\nabla^d X_t$ .

The ARIMA model (Autoregressive Integrated Moving Average Model) is a classical method for time series analysis and forecasting. Its fundamentals involve three key concepts: autoregression (AR), integration (I), and moving average (MA)[10].

Autoregression (AR): the autoregression component represents the relationship between the observations at the current moment and the observations at previous moments. This relationship can be expressed in terms of the autoregressive order ( $p$ ), which indicates how many lagged values are considered.

Integration (I): the integration is done to transform the original time series into a smooth time series. The integration order ( $d$ ) indicates how many times the time series needs to be differenced to smooth it. Usually, the first differencing ( $d=1$ ) is sufficient to smooth the time series, but in some cases more differencing may be required.

Moving Average (MA): is to capture short-term fluctuations in a time series by considering prior errors.

The ARIMA model building consists of the following steps:

Step 1: Data preprocessing. Includes checking the smoothness of the data and differencing if it is not smooth until a smooth sequence is obtained.

Step 2: Model identification. Select appropriate  $p, d, q$  values by looking at the plots of autocorrelation function (ACF) and partial autocorrelation function (PACF).

Step 3: Parameter Estimation. Estimate the parameters in the model using methods such as maximum likelihood estimation.

Step 4: Model fitting. Fit the data by taking the estimated parameters into the ARIMA model.

Step 5: Model Diagnosis. Check whether the residuals of the model are white noise, if not, the model needs to be readjusted.

Step 6: Prediction. Use the fitted ARIMA model for prediction of future values.

### 2.3. Establishment of profit model

In the replenishment and sales process, cost = replenishment volume  $\times$  wholesale price, i.e., cost can be expressed as  $v_{it}R_i$ .

For the sales volume, we can get a limiting factor for the sales volume using the pricing-volume model. Meanwhile, the demand predicted with the ARIMA model is the ideal state in the transaction, so the actual sales volume should be less than the demand. Naturally, the sales volume will also be less than the replenishment volume if the backorder business model is not considered. Thus, we take the actual sales volume as the minimum of the predicted sales volume, demand and replenishment volume, which is  $\min\{M_{it}, W_{it}, v_{it}\}$ . Further, defining the total income as  $F_{it}$ , we can obtain the income model as:

$$F_{it} = \min\{M_{it}, W_{it}, v_{it}\} \times T_{it} \quad (9)$$

Assuming that the profit is  $Q_{it}$ , based on the relationship between profit and cost and revenue, we can build the maximising revenue model as in equation (10):

$$\max \sum Q_{it} = \max \sum_{t=1}^7 \sum_i^6 (F_{it} - v_{it}R_i), \quad 1 \leq i \leq 6, \quad 1 \leq t \leq 7 \quad (10)$$

$$\begin{cases} R_i \geq 0 \\ v_{it} \geq 0 \\ h_{it} \geq 0 \end{cases} \quad (11)$$

In the actual sales process, the wholesale price  $R_i$ , the replenishment  $v_{it}$  should be greater than 0 in order to be meaningful, and the superstore wants to make a profit, the most basic to ensure that the pricing is greater than the cost, that is, the markup rate  $h_{it}$  should also be greater than 0. Therefore, it results in the constraints of (11).

### 3. Results

#### 3.1. Cost plus pricing method - establishment of pricing sales model

Cost-plus pricing is a method of setting the price of a product based on the unit cost of the product plus a percentage of the profit. Most companies determine the size of the profit added by the cost profit string. That is: price = unit cost + unit cost x cost margin = unit cost (1 + cost margin). Suppose  $h_{it}$  is the cost margin and  $T_{it}$  is pricing, and so:

$$T_{it} = (1 + h_{it})R_i \quad (12)$$

The pricing-volume model is built in terms of categories, while the sales unit price and wholesale cost in the data are in terms of individual items, so the price of individual items is used to determine the price of the category, and the key to this process is the determination of the weight of each individual item. Define  $k_i (\sum k_i = 1)$  as the weight of individual item  $i$  in the category, and the weight is the ratio of the total sales volume of the individual item to the total sales volume of the category. Suppose the wholesale price of a single item is  $m_i$ , the wholesale price of a category is  $R_i$ , and  $i$  denotes the type of category. Thus, the cost of the category is expressed as:

$$R_i = \sum k_i m_i \quad (13)$$

Substituting the data yields the weighted cost  $R_i$  for the category as shown in the table 1.

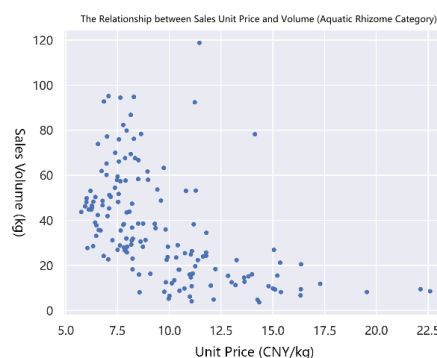
**Table.1.** Cost of categories  $R_i$

| category | Flower leaf | Chili peppers | edible fungi | florescent vegetables | Aquatic rhizomes | Solanaceae |
|----------|-------------|---------------|--------------|-----------------------|------------------|------------|
| $R_i$    | 4.193       | 6.614         | 9.745        | 4.294                 | 6.436            | 5.826      |

Table 1 shows the costs of different categories. These cost values indicate the cost of production for the corresponding category, which may include raw materials, labor in the production process and other related costs. These cost values can be used to compare the production cost of each category and help superstores make appropriate decisions in pricing and replenishment.

Similarly, based on the weighting method, using the sales price of individual items, we can find the historical pricing  $\alpha_i$  for each category.

Observe the historical total sales and historical pricing of each category and plot a scatterplot, as in Figure 1.



**Fig. 1** Observation - fitting of the relationship between sales unit price and sales volume (aquatic rhizomes)

It is clear from Figure 1 that the relationship between the unit price of sales and sales volume conforms to a logarithmic relationship, so the logarithmic function is considered fitting. Assume that the sales volume is  $W_i$ , so the relationship between sales volume and selling price is satisfied:

$$W_i = a_i + b_i \ln \alpha_i \quad (14)$$

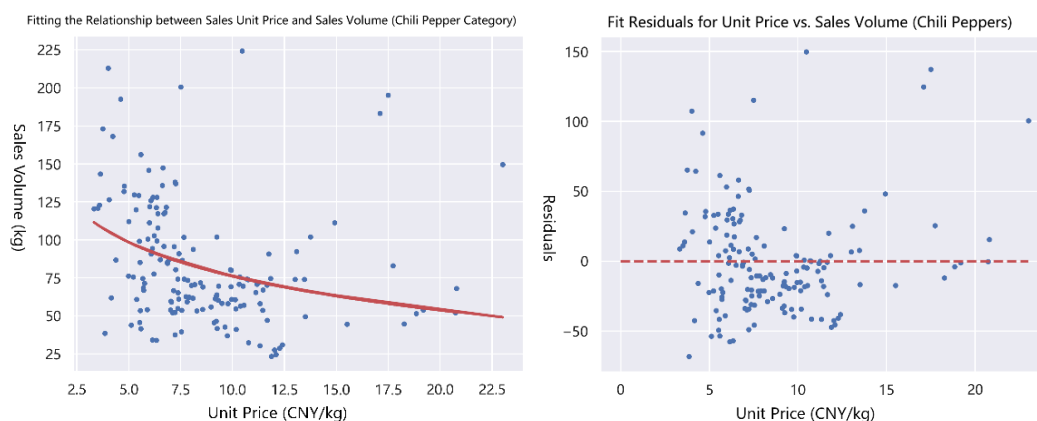
Further, for the coming week (1-7 July 2023), assuming that the sales volume is  $W_{it}$  and the pricing is  $T_{it}$  (where  $i$  denotes the category,  $1 \leq i \leq 6$ ;  $t$  denotes the  $t$ th day of July  $1 \leq t \leq 7$ ), the relationship between the sales volume and the selling price is satisfied:

$$W_{it} = a_i + b_i \ln T_{it} \quad (15)$$

The parameters  $a_i$  and  $b_i$  corresponding to the L-M fitting algorithm for each category using historical sales flow data are shown in Table 2. The fitting effect of the excerpted chilli pepper categories is shown in Figure 2.

**Table.2.** Parameters of each category

| category | Flower leaf | Chili peppers | edible fungi | florescent vegetables | Aquatic rhizomes | Solanaceae |
|----------|-------------|---------------|--------------|-----------------------|------------------|------------|
| $a_i$    | 316.041     | 88.436        | 130.007      | 46.461                | 150.351          | 154.320    |
| $b_i$    | -80.278     | -35.859       | -50.901      | -22.911               | -70.287          | -70.429    |



**Fig. 2** Fitting the Relationship between Sales Unit Price and Sales Volume (Excerpt)

After the model building and solving process, we evaluate our pricing-sales sub-model by observing the fitted residual plots and calculating the goodness-of-fit  $R^2$

Observing the scatter plot of the residuals after fitting for each category, it can be found that the data points are more uniformly distributed in the vicinity of the horizontal axis  $y=0$ , and there is no obvious bias in the fitting. Calculate the goodness of fit  $R^2$ :

$$R^2 = 1 - \left( \frac{SSR}{SST} \right) \quad (16)$$

The  $R^2$  results for each category were calculated as in Table 3.

**Table.3.** Each category  $R^2$

| category | Flower leaf | Chili peppers | edible fungi | florescent vegetables | Aquatic rhizomes | Solanaceae |
|----------|-------------|---------------|--------------|-----------------------|------------------|------------|
| $R^2$    | 0.107       | 0.122         | 0.267        | 0.098                 | 0.101            | 0.084      |

$R^2$  is significantly lower for each category. Combining observations of the original data with the fitted curves, we believe that this is caused by fluctuations in sales data within a similar selling price range due to a large number of sales data and a wide range of influencing factors. By analyzing the

fitted curves with the residuals, we conclude that the fit is still a good approximation of the centroid of the distribution of sales at the same selling price.

### 3.2. Category demand forecast results

From the supply-side relationship, although the demand is characterized by elasticity in the macro-market economic environment, it has a general trend. In this regard, the ARIMA time series was used to construct a model to forecast the future demand for the category based on the sales volume flow data for each July of history, and June 2023, using the ARIMA time series.

Substituting the data to solve the problem yields the results shown in Figure 3. Due to space constraints, we chose chilli as our representatives. From the figure, it can be seen that the predicted results are similar to the recent fluctuation frequency of the time series, indicating that the model has short-term memory, which is also in line with the historical law of demand in the vegetable market, indirectly demonstrating the rationality of the model.

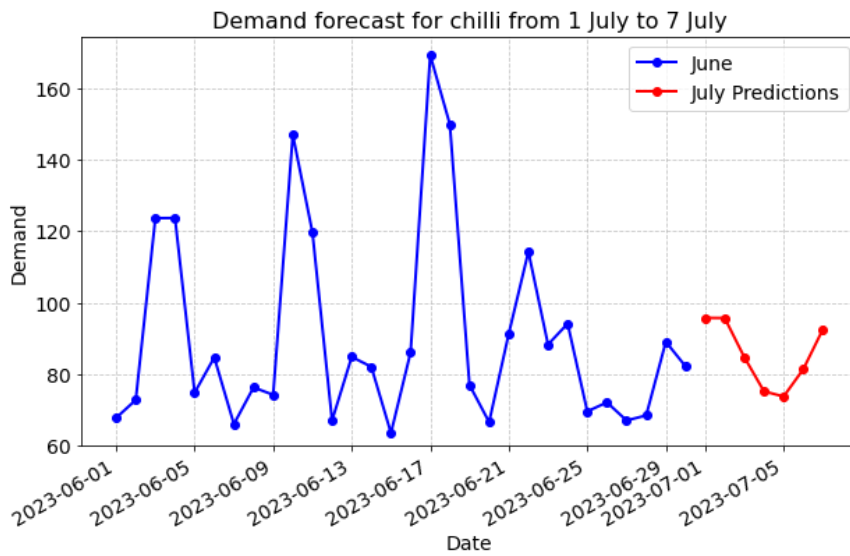


Fig. 3 Forecast of demand for each category (excerpt)

### 3.3. Results of profit model

For the objective function

$$\sum_{t=1}^7 \sum_{i=1}^6 Q_{it} \quad (17)$$

Since  $Q_{it}$  is independent for any  $i, t$  that satisfies the condition, we only need to consider maximizing  $Q_{it}$  for each  $i, t$ . In fact,

$$Q_{it} = F_{it} - v_{it}R_i \quad (18)$$

Since the objective function contains a minimum value, we consider a categorical discussion.

#### ● Situation I

When  $\min\{M_{it}, W_i, v_{it}\} = M_{it}$ , The model obtains further constraints, i.e.

$$\begin{cases} M_{it} \leq W_i = a_i + b_i \ln[(1 + h_{it})R_i] \\ M_{it} \leq v_{it} \end{cases} \quad (19)$$

From the first restriction of (19)

$$h_{it} \leq \frac{1}{R_i} \exp\left(\frac{M_{it} - a_i}{b_i}\right) - 1 \quad (20)$$

And the objective function can be written as

$$Q_{it} = M_{it}T_{it} - v_{it}R_i = R_i(M_{it} + M_{it}h_{it} - v_{it}) \quad (21)$$

Since by (20) we have that the maximum value of  $h_{it}$  is independent of  $v_{it}$ , the maximum value of  $h_{it}$  will be equal to the maximum value of  $v_{it}$  when.

$$h_{it} = \frac{1}{R_i} \exp\left(\frac{M_{it} - a_i}{b_i}\right) - 1, v_{it} = M_{it} \quad (22)$$

Then the objective function will take the maximum value:

$$Q_{max,I} = R_i M_{it} \left( \exp\left(\frac{M_{it} - a_i}{b_i}\right) - 1 \right) \quad (23)$$

### ● Situation II

When  $\min\{M_{it}, W_i, v_{it}\} = W_i = a_i + b_i \ln T_{it}$ , The model obtains further constraints.

$$\begin{cases} v_{it} \geq W_i = a_i + b_i \ln T_{it} \\ M_{it} \geq W_i \end{cases} \quad (24)$$

The objective function is then rewritten as

$$\begin{aligned} Q_{it} &= (a_i + b_i \ln T_{it})T_{it} - v_{it}R_i \\ &= R_i[a_i + b_i \ln[(1 + h_{it})R_i] + h_{it}a_i + h_{it} \ln[(1 + h_{it})R_i] - v_{it}] \end{aligned} \quad (25)$$

Bringing in the minimum value of  $v_{it}$ , we have

$$Q_{it} \leq h_{it}a_i + h_{it}b_i \ln[(1 + h_{it})R_i] \quad (26)$$

Deriving from  $h_{it}$  on the right-hand side, we obtain

$$a_i + b_i \ln[(1 + x)c] + \frac{b_i x}{1 + x} \quad (27)$$

It is a typical transcendental equation for which no analytical solution can be obtained, however it is easily maximized by numerical methods. Thus,  $Q_{max,II}$  is denoted as the numerical maximum of the function.

$$f(h_{it}) = h_{it}a_i + h_{it}b_i \ln[(1 + h_{it})R_i] \quad (28)$$

### ● Situation III

When  $\min\{M_{it}, W_i, v_{it}\} = v_{it}$ , the model obtains further constraints.

$$\begin{cases} v_{it} \leq M_{it} \\ v_{it} \leq a_i + b_i \ln[(1 + h_{it})R_i] \end{cases} \quad (29)$$

The objective function is expressed as

$$Q_{it} = v_{it}(1 + h_{it})R_i - v_{it}R_i \quad (30)$$

From the first constraint of (2), one can further obtain a constraint on  $h_{it}$

$$h_{it} \leq \frac{1}{R_i} \left( \exp\left(\frac{v_{it} - a_i}{b_i}\right) - 1 \right) \quad (31)$$

Thus, the objective function can be immediately rewritten as

$$Q_{it} = v_{it} \left( \exp\left(\frac{v_{it} - a_i}{b_i}\right) - 1 \right), v_{it} \leq M_{it} \quad (32)$$

Considering that the derivatives of this function are also transcendental equations and have constraints, the optimization is still carried out numerically and the result is recorded as

$$Q_{max,III} = \max \left\{ v_{it} \left( \exp \left( \frac{v_i - a_i}{b_i} \right) - 1 \right) \right\}, v_{it} \leq M_{it} \quad (33)$$

After obtaining the maximum value for each of the three cases, we consider further obtaining the global maximum value of the objective function.

For each set  $i, t$  due to the presence of two decision variables, we use Python to search in the feasible space of  $h_{it}, v_{it}$  and attribute each pair  $(h_{it}, v_{it})$  to one of the three scenarios, and finally obtain a set  $(h_{it0}, v_{it0})$  that maximizes the objective function  $Q_{it}$ . Finally, iterate  $i=1, 2, \dots, 6$ ;  $t=1, 2, 7$  to obtain the 7-day profit maximum  $\sum_{t=1}^7 \sum_{i=1}^6 Q_{it}$  with the optimal markup rate  $h_{it}$  and replenishment  $v_{it}$  per category per day.

Substituting the data we solved for the following results (Table 4-5):

**Table.4.** Category number

| $i$      | 1           | 2             | 3            | 4                     | 5                | 6          |
|----------|-------------|---------------|--------------|-----------------------|------------------|------------|
| category | Flower leaf | Chili peppers | edible fungi | florescent vegetables | Aquatic rhizomes | Solanaceae |

Maximum profit for:

**Table.5.** Maximum profit

| $t \backslash i$ | 1       | 2       | 3       | 4       | 5       | 6       | 7       |
|------------------|---------|---------|---------|---------|---------|---------|---------|
| 1                | 276.301 | 255.782 | 227.259 | 202.075 | 205.721 | 235.067 | 271.472 |
| 2                | 23.003  | 23.048  | 23.021  | 23.016  | 22.931  | 23.016  | 23.009  |
| 3                | 10.102  | 10.015  | 10.021  | 10.149  | 10.026  | 10.059  | 10.059  |
| 4                | 9.013   | 9.025   | 9.053   | 9.076   | 9.025   | 9.007   | 9.025   |
| 5                | 9.207   | 9.209   | 9.208   | 9.270   | 9.154   | 9.264   | 9.207   |
| 6                | 20.947  | 20.998  | 20.988  | 20.974  | 20.998  | 20.988  | 20.888  |

Therefore, based on this replenishment pricing strategy, the superstore's profit in the coming week (1-7 July 2023) is maximized as  $\max \sum Q_{it} = 2179.677$ (yuan).

## 4. Conclusions

In this paper, the focus is on studying the vegetable category. The cost of the category is determined using sales volume as the weight. The relationship between sales volume and pricing is fitted using the L-M algorithm. The ARIMA algorithm is employed to predict demand. A revenue maximization model is constructed by selecting the minimum of demand, sales volume, and replenishment as the predicted sales volume. Numerical methods are utilized to solve for the optimal solution. Experimental results indicate that the restocking volume for the Flower leaf category is the highest, and the profit for the upcoming week surpasses that of other categories. In addition, based on the pricing and replenishment strategy derived from the model in this paper, the superstore gets a maximum profit of 2179.68 yuan in the coming week. However, the mathematical model developed in this paper does not fully consider the potential impact of sudden catastrophic weather events on the profitability of superstores. Therefore, future research should aim to comprehensively incorporate more realistic factors to enhance the practicality and interpretability of the constructed mathematical model.

## References

- [1] Hu Yuanyuan. Research on the current situation and countermeasures of cold chain logistics for fresh agricultural products[J]. China Navigation Weekly, 2023(47):160-162.
- [2] Wei Feng, Guo Xiaofei. A review of demand forecasting methods for fresh agricultural products[J]. Logistics Technology, 2024,47(01):29-31.
- [3] Li Yali. Joint replenishment and pricing coordination decision considering investment in preservation technology[D]. Chongqing: Chongqing Jiaotong University, 2022.
- [4] Liang Aonan, Wang Shuyun. Study on two-stage sales pricing and production optimization of fresh agricultural products[J]. Highway Traffic Science and Technology, 2023, 40(06):225-233.
- [5] Lu Zhigang, Chen Cheng. Pricing strategy of fresh agricultural products retailers based on customer classification[J]. Logistics Technology, 2023, 42(11):35-42.
- [6] Hao Yangyang, Zou Yu. Shanghai fresh agricultural products logistics demand forecasting based on BP neural network[J/OL]. Journal of Shanghai Maritime University:1-8[2024-01-21].
- [7] Yao Chun, Lü Huiyan. Pricing strategy of fresh agricultural products under different loss and discount factors[J]. Modern Business, 2020(26):11-12.
- [8] Han X, Bai S, Yu H. Pre-sale Strategy for Fresh Agricultural Products under the Consideration of Logistics Services[J]. IEEE Access, 2021, 9: 127742-127752.
- [9] Li Jia, Qiu Wei, Zou Shaofen. Research on Magnetic Target Localization Method Based on Improved LM-GN Algorithm[J]. Digital Ocean and Underwater Attack and Defense, 2023, 6(05):552-561.
- [10] Zou Xiaojiang, Zhao Han, Wang Qiyin, et al. Research on the Forecast of Influenza in Chongqing City Based on the ARIMA Model[J]. Journal of Chongqing Medical University, 2023, 48(12):1425-1429.