# Research on replenishment decision of vegetable commodities based on genetic algorithm

Ruohan Feng[1, *, #], Huacheng Zhu[2, #], Luyao Zhang[1] and Yiping Song[1]

[1]College of Physics Programs and Technology, Shenyang Normal University, Shenyang, China

[2]School of Mathematics and Systems Science, Shenyang Normal University, Shenyang, China

*Corresponding author: 3394344206@qq.com

#These authors contributed equally to this work

**Abstract.** Fresh produce superstores, are an important component of the modern economy, however, they face multiple challenges in vegetable merchandise management, including how to forecast accurate sales trends, develop reasonable pricing strategies, and effective replenishment programs. Therefore, this study aims to provide a comprehensive set of vegetable merchandising strategies for superstores to help them better solve their problems. Firstly, this paper constructs an integrated fitting model to fit a polynomial function to each single vegetable item; secondly, taking amaranth as an example, it constructs an image of the parameters of the amaranth fitting function; based on the ARIMA model of various categories of vegetable demand and wholesale price prediction, it solves the amaranth demand using the ARIMA model, and uses the hyper-parameter optimization to obtain the model with the smallest BIC value; it establishes the ARIMA (5,1,6) model and forecast the demand for the next seven periods, with the objective of maximizing the revenue of the superstore, establish a linear programming model (LP), constraints, and solve the model based on genetic algorithm.

**Keywords:** Linear Programming Model; ARIMA; Genetic Algorithm; Vegetable Commodity Replenishment.

## 1. Introduction

In a fresh food superstore, vegetables have a short shelf life and their quality deteriorates with selling time, and most of them cannot be re-sold the next day if they are not sold on the same day. Therefore, supermarkets replenish their stocks daily based on historical sales and demand. However, because supermarkets sell a variety of vegetables of different origins, and because the buying and trading hours are usually from 3:00 a.m. to 4:00 a.m., merchants need to make replenishment decisions without knowing the specific varieties and prices of the vegetables they buy. At the same time, the pricing of vegetables generally adopts the "cost-plus pricing" method, and supermarkets will sell goods with reduced quality at a discount [1].

In this paper, we consider a category-based replenishment plan for supermarkets, analyze the relationship between the total sales volume of each vegetable category and cost-plus pricing, and provide daily replenishment volume and pricing strategies for July 1-7, 2023, with the goal of maximizing the revenue of supermarkets [2].

## 2. An integrated model of the relationship between total sales and cost-plus pricing

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. It is also a method of adding a portion of profit to the cost of the product. Most companies are based on the cost margin to determine the size of the added profit. That is: price = unit cost + unit cost × cost margin = unit cost (1 + cost margin).

In this paper, exploring the relationship between total daily sales and cost-plus pricing can be viewed as exploring the relationship between total daily sales and cost margin.

In order to depict the relationship between total daily sales and cost-plus pricing, an integrated fitting model is constructed to fit a polynomial function to each individual vegetable item. The single fitting model can be formulated as:

$$y = f(x, \theta_0, \theta_1, \cdots, \theta_p) + \varepsilon_i \tag{1}$$

Where y is the daily sales volume; x is the cost margin; $\theta_0, \theta_1, \cdots, \theta_p$ are p+1 parameters to be estimated; and $\varepsilon_i$ are random variables independent of each other and obeying the same normal distribution $N(0, \sigma^2)$.

In previous studies, polynomials of 1 to 5 times are commonly used to fit the data. Different datasets fitted with different times of polynomial times are used to fit different data sets with different numbers of polynomials. In order to maximize the accuracy of the model constructed in this paper, an integrated fitting model is constructed to represent the relationship between total daily sales and cost-plus pricing. The specific approach is as follows:

$$F(x) = max\{f_1(x)_{r^2}, f_2(x)_{r^2}, f_3(x)_{r^2}, \ldots, f_5(x)_{r^2}\} \tag{2}$$

For each individual product the model with the best results from 1 to 5 fits was selected as the fitting function. To maximize the accuracy of the model.

The amaranth was fitted as an example. The resulting model is shown below:

$$F(x) = \theta_1 x^5 + \theta_2 x^4 + \theta_3 x^3 + \theta_4 x^2 + \theta_5 x + \theta_6 \tag{3}$$

$$\theta_1 = 2.891400660730282e + 02 \tag{4}$$

$$\theta_2 = -1151.16934122665 \tag{5}$$

$$\theta_3 = 1630.20794866095 \tag{6}$$

$$\theta_4 = -1019.37017030623 \tag{7}$$

$$\theta_5 = 277.745717273839 \tag{8}$$

$$\theta_6 = -17.1900175697818 \tag{9}$$

The fitted images of their 1-5 times functions are plotted as shown in Figure 1:

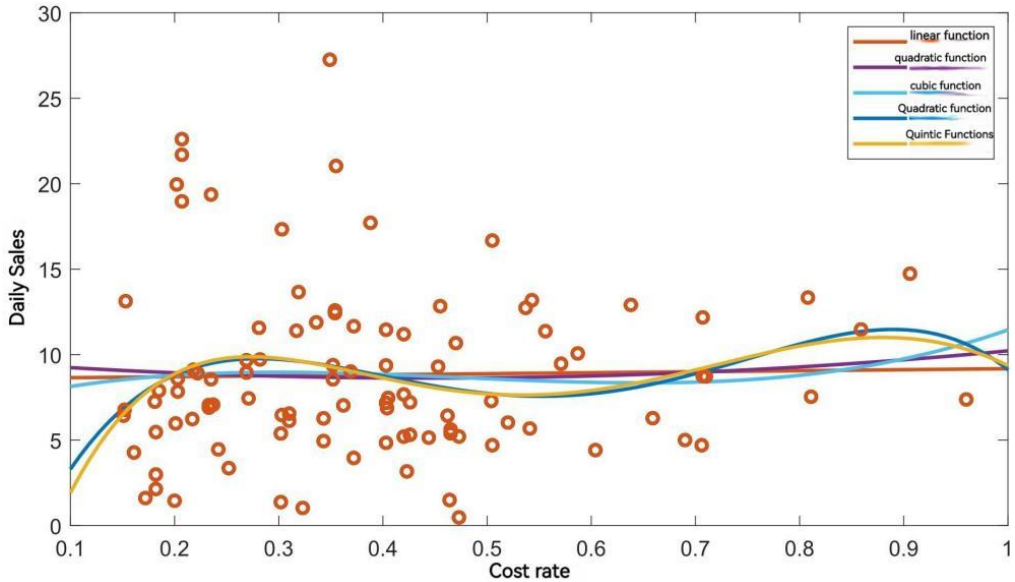


**Fig. 1** The function fits the graph

# 3. Demand and Wholesale Price Forecasts for Various Types of Vegetables Based on ARIMA Models

## 3.1. Predictive modeling

Reliable market demand analysis is particularly important for replenishment and pricing decisions. If future market demand for various types of vegetables can be forecasted, it will provide guidance for replenishment decisions and pricing decisions.

Traditional time-series forecasting methods are mainly used to forecast time series through statistical and mathematical models, including the early exponential smoothing method, AR, MA, and ARMA models for forecasting linear, univariate, and smooth time series. The method often does not take into account its stability requirements, resulting in its low applicability to the data. The following equation is the defining expression for the ARMA model [3].

$$x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + \varepsilon_t + \sum_{j=1}^{q} \vartheta_j \varepsilon_{t-j} \tag{10}$$

Where ε is the white noise sequence, p and q are the orders, $\varnothing_j$ refers to the AR coefficients, and $\theta_j$ refers to the MA coefficients. When q = 0, the model is an AR model, which indicates that the current data at moment t is affected by the previous p moments; when p = 0, the model is an MA model, which indicates that the current data at moment t is related to the random disturbance terms ( $\varepsilon_t$, $\varepsilon_{t-1}$, $\varepsilon_{t-2}$, ···) in the previous q moments, but not to the data in the previous q moments. p-order AR model The p-order AR model and the q-order MA model together constitute the ARMA model for predicting smooth series.

The essence of the ARIMA model is the combination of the difference operation with the ARMA model, denoted as ARIMA(p,d,q).The ARIMA model can be formulated as:

$$\Phi(B)(1-B)^d y_t = e(B)\varepsilon_t \tag{11}$$

Where $y_t$ is the time series to be predicted, d is the difference order, p is the number of autoregressive terms, q is the number of moving average terms, $\varepsilon_t$ is the zero-mean white noise sequence, and B is the delay operator. b satisfies the following expression:

$$B^n y_t = y_{t-n} \tag{12}$$

$$\Phi(B) = 1 - \Phi_1 B - \cdots - \Phi_p B^p \tag{13}$$

$$\theta(\mathrm{B}) = 1 - \theta_1 B - \cdots - \theta_q B^q \tag{14}$$

The selection of parameters (p,d,q) has a large impact on the predictive performance of the ARIMA model. The parameter d can be determined during the differencing process and d is numerically equal to the order of the differencing.

In the traditional ARIMA study, many scholars determine the parameters of p,q through ACF and PACF. However, there is no quantitative selection criterion for selecting parameters using this method, and it relies on the subjective experience of the researcher to a certain extent. Moreover, in the era of big data, this method cannot be applied to large-scale and frequent prediction. In order to enhance the practicality of this model, this paper adopts Bayesian Information Criterion (BIC) to select p,q, which balances the model's goodness-of-fit to the data and the model's complexity. BIC can be formulated as follows:

$$BIC = -2log\,(\hat{L})A + klog\,(n) \tag{15}$$

Where $\hat{L}$ is the maximum value of the likelihood function. n is the number of samples.

## 3.2. ARIMA model solving

The demand for amaranth is still used as an example for solving the ARIMA model. Plot the original series as shown in Figure 2:

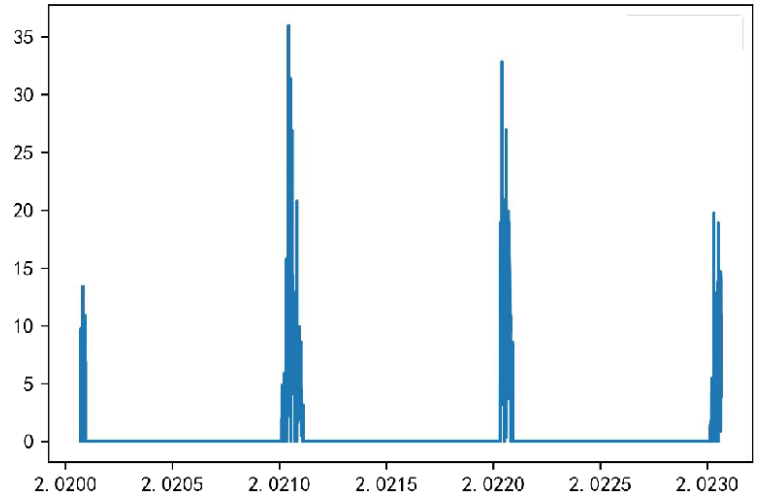


**Fig. 2** Original sequence diagram

Based on the image, we can see that the sequence is periodic and shows a certain trend, so it is first-order differenced to obtain a differential sequence plot as in Figure 3:

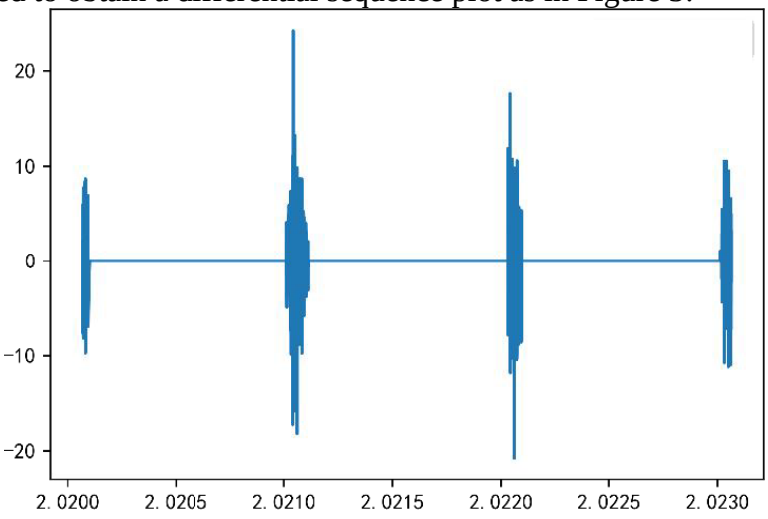


**Fig. 3** Difference chart

It can be observed that at this point the series no longer contains the trend term at all. The autocorrelation and partial autocorrelation after differencing are plotted as shown in Figure 4:

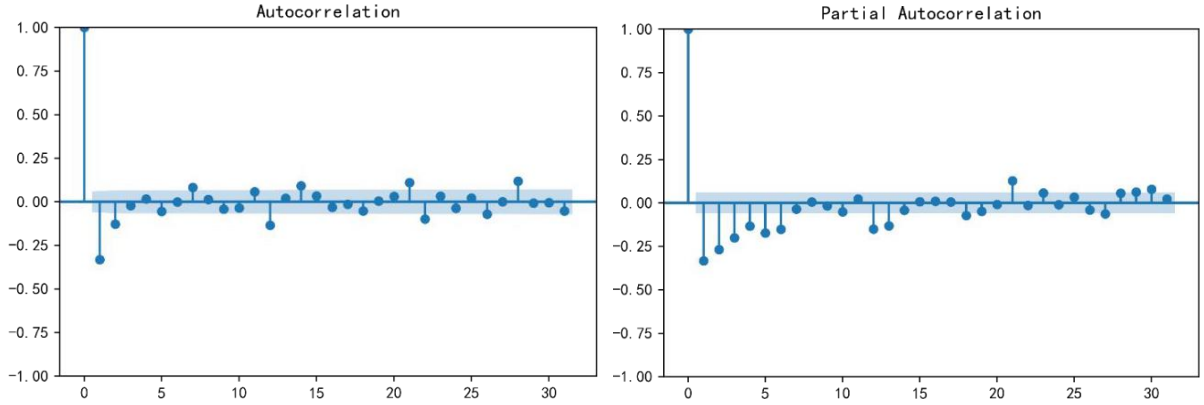


**Fig. 4** Autocorrelation (left) and bias correlation (right) plots after differencing

The model with the smallest BIC value is obtained using hyperparameter optimization and the BIC matrix is obtained as shown in Table 1:

**Table 1**. Three Scheme comparing

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 0 | 5949.621 | 5676.450 | 5636.008876 | 5641.907515 | 5648.849548 | 5655.515001 | 5660.933106 |
| 1 | 5829.833 | 5636.106 | 5641.967229 | 5648.608384 | 5655.857274 | 5662.226315 | 5667.822655 |
| 2 | 5756.098 | 5641.930 | 5648.916737 | 5655.596678 | 5655.271629 | 5654.11471 | 5665.949639 |
| 3 | 5718.958 | 5648.917 | 5655.815482 | 5649.005054 | 5662.022513 | 5662.134696 | 5674.197587 |
| 4 | 5706.701 | 5655.904 | 5662.654889 | 5660.801542 | 5667.250947 | 5654.957065 | 5642.977231 |
| 5 | 5681.410 | 5661.331 | 5668.022036 | 5660.421837 | 5673.172687 | 5675.399838 | **5635.334276** |
| 6 | 5663.289 | 5669.071 | 5675.30625 | 5667.087542 | 5674.105997 | 5676.195251 | 5640.05495 |

In the BIC matrix, the rows represent the values of p and the columns represent the values of q. 5635.334276 which minimizes the loss of information is selected, i.e., the optimal parameter combination is p=5,q=6.

Based on the above analysis, we build ARIMA (5,1,6) model and forecast the demand for the next 7 periods. The results are shown in Table 2:

**Table 2**. Projected value

| Dates | Projected value |
|---|---|
| 20230701 | 10.592789 |
| 20230702 | 10.420175 |
| 20230703 | 7.430967 |
| 20230704 | 6.524542 |
| 20230705 | 5.717339 |
| 20230706 | 8.03295 |
| 20230707 | 9.6711 |

## 4. Maximize superstore revenue

### 4.1. Planning modeling

A linear programming model (LP) is developed with the objective of maximizing superstore revenue. The objective of the model is:

$$max \sum_t \sum_k P_t = A_{k,t}(S_{k,t} - C_{k,t}) \tag{16}$$

Equation (16) is the optimization objective, i.e., each vegetable category maximizes the revenue of the superstore in the coming week. Where $P_t$ is the revenue of the superstore on day t; $C_{k,t}$ is the wholesale price per kilogram of vegetable k on day t; $S_{k,t}$ is the selling price per kilogram of vegetable k on day t; and $A_{k,t}$ is the sales volume of vegetable k on day t.

The constraints of the model are as follows:

$$Im_{k,t} \geq A_{k,t} \tag{17}$$

$$\hat{Y}_{k,t} > Im_{k,t} \tag{18}$$

$$Im_{k,t} \geq 0 \tag{19}$$

$$A_{k,t} = 0_{k,t} - I_{k,t} \tag{20}$$

$$S_{k,t} = C_{k,t}(1 + p_{k,t}) \tag{21}$$

Equation (17) indicates that the sales volume of dish k on day t must not be greater than the corresponding replenishment volume. Equation (18) indicates that the replenishment quantity of dish k on day t is less than the predicted demand. Equation (19) indicates that the amount of incoming

stock is a non-negative number. Equation (20) indicates that the sales volume of dish k on day t is equal to the actual outflow minus the return volume. Equation (21) is the formula for cost-plus pricing.

### 4.2. LP solving based on genetic algorithm

Genetic algorithm, as one of the most commonly used intelligent optimization algorithms, has been developed and used by many scholars for its excellent exploration and exploration performance. Genetic algorithm imitates the evolutionary process of population in biology when performing optimization, and explores the solution space under the premise of retaining the excellent individuals and gene fragments in the population through the two operators of crossover and mutation. The crossover operator accomplishes the exchange and transmission of good gene segments among individuals in the population, and the mutation operator accomplishes the exploration of the solution space to avoid falling into a Local optimization [4].

In this paper, a separate scheme is designed for each day, i.e., each run gets the daily replenishment total and pricing strategy for the day. The next section will focus on the coding, crossover, mutation and decoding schemes designed in this paper.

The coding scheme is shown in Figure 5:




**Fig. 5** Coding scheme

In this case, the chromosome length is fixed and represents the total number of categories of vegetables. Each gene is fixed to represent the profitability of one category, and the meanings represented by the values at the fixed loci do not change with changes in the lower corner markers.

In crossover, we randomly select two points as mutation sites and then exchange the corresponding positions of the two sequences. The process is shown in Figure 6:

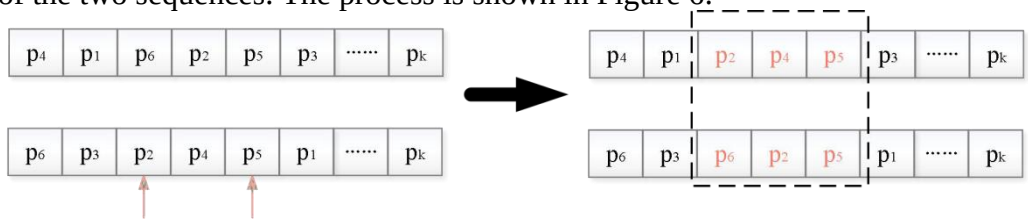


**Fig. 6** Coding scheme

During mutation, a random point is selected for random mutation.

The daily pricing strategy for each category of vegetables is easily obtained by solving the chromosome inversely. Further, the relationship between total sales and cost-plus pricing has been solved, at which point the cost margin can be obtained to invert the predicted replenishment for that day by using the discount rate, calculated as shown in the following equation.

$$Im_{k,t} = \frac{A_{k,t}}{1-impair_k} \tag{22}$$

And replenishment is based on sales volume. Thus, the replenishment decision and the pricing decision are accomplished simultaneously. The genetic algorithm parameters are set as shown in Table 3 [5]:

**Table 3**. Genetic algorithm parameters

| Parameters | Retrieve value |
|---|---|
| Population size | 200 |
| Crossover rate | 0.7 |
| Variation rate | 0.3 |
| Number of iterations | 50 |

The convergence curve is plotted by taking the average value in each population. The convergence curve for seven days is shown in Figure 7:

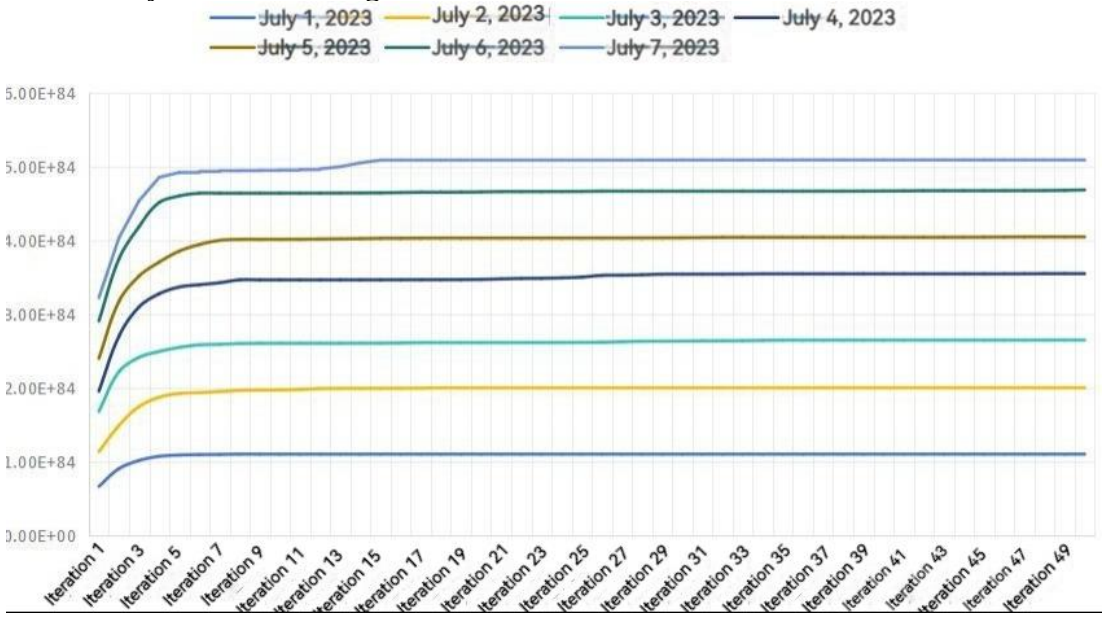


**Fig. 7** Convergence curve

From the figure, it can be noticed that the curve converges quickly in the first half of the curve and smoother in the second half, and when the number of iterations reaches about 15, the objective value has converged.

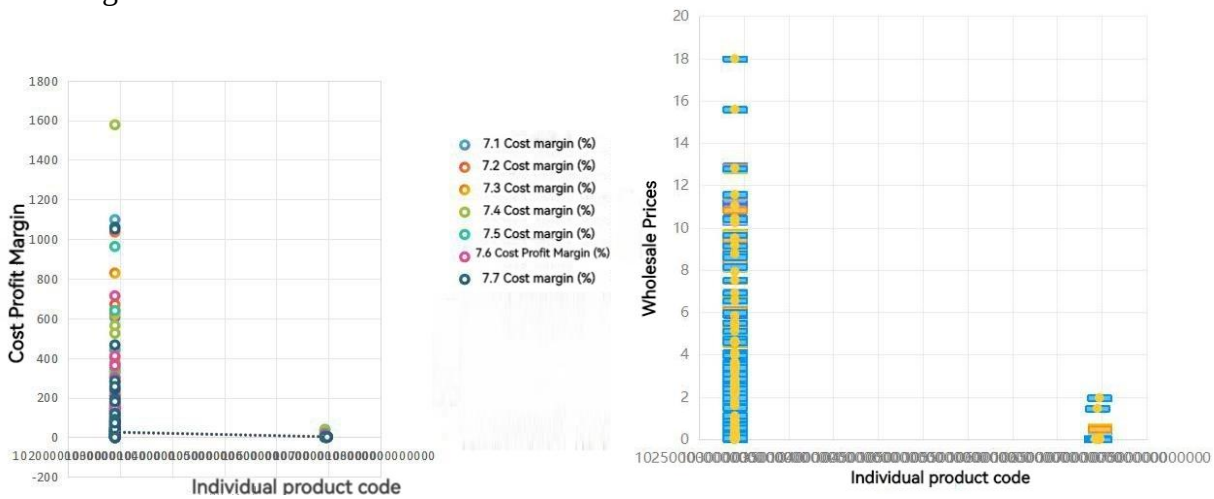


**Fig. 8** Pricing strategy

## 5. Summary

In this paper, based on the historical sales data, Matlab's fitting toolbox is utilized to establish the expression of the functional relationship between sales volume and cost-plus pricing. Then, an optimization model is constructed to solve the optimal replenishment plan and pricing strategy by taking the total daily replenishment and pricing strategy as the decision variables and the superstore revenue as the objective function, and considering the constraints of inventory management and sales demand.

This paper integrates various aspects of data processing, statistical analysis and optimization modeling to make it suitable for solving complex superstore operation problems. It also considers data quality, actual business requirements, and the utilization of additional data, which will help improve the accuracy and practicality of the model. However, it is important to ensure that these plans

can be effectively implemented during the actual modeling process and that practical decisions can be made based on the model results.

## References

[1] Lu Yajie.Research on dynamic pricing of quality fresh vegetables in supermarkets in China [D]. Beijing Jiaotong University, 2010.

[2] Gu Sihong. A study on dynamic pricing of fresh products in H-retailers considering freshness change [D]. Donghua University, 2023.

[3] Li N, & Wang QH. (2016). Time series analysis and its R language implementation. Mechanical Industry Press.

[4] Xiaodi Guan, Hongchao Guo, Hongchen Wang, et al. Research on optimization of TMD parameters based on genetic algorithm [J]. Journal of Xi'an University of Architecture and Technology (Natural Science Edition),2023,55(2):211-216.

[5] Gao Xiang, Ocean. Application of genetic algorithm [J]. Journal of Chifeng College (Natural Science Edition), 2009, 25(3): 21-22.