

Research on Vegetable Replenishment and Pricing Based on Cost-Plus Pricing Method and ARIMA Model

Xuanying Han¹, Ranran Zhang^{2,*}, Ziheng Li³

¹ Department of Mathematics, Tianjin University of Finance and Economics, Tianjin, China, 10070

² Department of Management Information System, Tianjin University of Finance and Economics, Tianjin China, 10070

³ Department of Computational Science and Data Technology, Tianjin University of Finance and Economics, Tianjin, China, 10070

* Corresponding Author Email: tufezrr@163.com

Abstract. Replenishment and pricing of vegetables are the focus of commercial supermarkets. Accurately predicting the future vegetable demand and pricing of commercial supermarkets is of great significance for commercial supermarkets to increase revenue. To make accurate predictions and maximize profits, this paper predicts the mean and variance of the demand and pricing of various vegetable categories in the future based on the cost-plus pricing method and the ARIMA time series model, and obtains the best demand and pricing for the next seven days. The results show that the model has passed the residual independent test and normal test, and R^2 is greater than 0.7, which has a good fitting effect. The model is suitable for the prediction of vegetable replenishment and pricing, and can provide a reference for vegetable management in supermarkets.

Keywords: Vegetable Replenishment, Vegetable Pricing, Cost-Plus Pricing, ARIMA Model.

1. Introduction

With the development of the economy and the improvement of living standards, people pay more attention to diet health. As an indispensable part of daily diet, vegetables have a huge market and their consumption patterns are becoming more and more complex. All kinds of vegetables sold in fresh supermarkets are vulnerable and perishable, with short shelf life and seasonal influence. Their appearance and quality will decrease with time, and most of the unsold varieties cannot be resold every other day [1-2]. To reduce costs and improve efficiency, supermarkets need to make daily replenishment and pricing based on the historical sales volume and price of vegetables, and formulate reasonable business strategies.

Scholars have carried out extensive discussions on the pricing and replenishment of fresh supermarkets. Sainathan A [3] studied the competitive pricing and replenishment of perishable products under dynamic demand substitution, but ignored the time-varying characteristics of historical data. Chen et al. [4] used a back propagation neural network and logistic regression to predict the sales of fresh food in convenience stores, but the model was poor. Fan T et al. [5] established a dynamic programming model for revenue management of fresh agricultural products considering consumer choice behavior, but did not consider holding costs and ordering costs. Bi Liyuan et al. [6], based on the prediction cost of recurrent neural networks, have a certain nonlinear mapping ability in prediction, but the overall ability is relatively weak. Most studies only consider the wholesale price of vegetables, and rarely consider the loss rate of vegetables, the fixed cost of sales (such as site fees, management fees, etc.) and other factors. In addition, although the machine learning method has been applied, for the time series dominated by linear components, the fitting effect is unstable and the accuracy needs to be improved. Dutta S et al. [7] predicted the price of onion and potato based on the ARIMA model, showing certain advantages.

Given the problems existing in the current research, this paper considers market supply and demand, fixed cost and vegetable loss rate. In this paper, the cost-plus pricing method is used to establish the pricing model and the replenishment model.

In the prediction process, this paper uses the ARIMA model to predict the wholesale price and daily sales volume, and judges the goodness of fit of the model through residual independent test and normal test. According to the characteristics of historical data changing with time, this paper uses the ARIMA model to predict the wholesale price and daily sales volume, and judges the goodness of fit of the model through the residual independent test and the normal test. Finally, the prediction results are brought into the pricing model and replenishment model to calculate the sales price and replenishment volume, to provide a reasonable business strategy for fresh supermarkets. The research in this paper helps to solve the problem of pricing and replenishment of fresh supermarkets, and reduces the loss rate of vegetables, which has achieved certain economic benefits.

2. Model Construction

This paper uses the cost-plus pricing method to price, uses the ARIMA model to predict the wholesale price and sales volume of each vegetable category in the future, and then uses the cost-plus pricing formula to obtain pricing and replenishment.

2.1. Cost-Plus Pricing Method

The cost-plus pricing method [8-10] is a method of setting product prices according to the unit cost of the product plus a certain proportion of profits. In actual production activities, supermarkets generally use this method to determine the price of vegetables, which is simple and easy to use, and can ensure that supermarkets can obtain certain benefits while returning. The calculation formula is shown in Formula (1).

$$P = C \cdot (1 + r) \quad (1)$$

Where P is the price, C is the cost, and r is the cost-plus rate.

A category contains multiple items, so the sales volume, wholesale price and loss rate of each item are different. Therefore, the proportion of each item should be taken into account when pricing in category. The linear weight method can assign a weight to each small index, and the large index covering comprehensive information can be obtained by using the weighted summation method. To price and replenish more comprehensively and accurately, this paper takes the proportion of daily sales of each single product in a certain category as the weight, and uses the linear weight method to calculate the wholesale price and loss rate of vegetable categories.

Category A contains vegetable items 1, 2..., n, and its sales volume is $\begin{bmatrix} s_{11} & \cdots & s_{1m} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nm} \end{bmatrix}$ in m days.

Calculate the weight of vegetable items and the daily wholesale price of vegetable categories:

$$w_{ij} = \frac{s_{ij}}{\sum_{j=1}^m s_{ij}} \quad (2)$$

$$p_j = \sum_{i=1}^n p_{ij} \cdot w_{ij} \quad (3)$$

Where i is the item number, j is the date number, w_{ij} is the weight of the item, p_{ij} is the daily wholesale price of the item, and p_j is the daily wholesale price of the category.

The loss rate is obtained according to the data of the recent inventory cycle. In this paper, the inventory cycle is set to nearly two months, and the proportion of daily sales of each item in a category in the past two months is selected as the weight. The linear weight method is also used to solve the loss rate of vegetable categories. Suppose that the total sales of n vegetable items in category A in the past two months are $sr_1, sr_2, \dots, sr_n$, respectively. Calculate the sales weight of vegetables in the past two months and the category loss rate in the past two months, as shown in Formula (4) and Formula (5).

$$wr_i = \frac{sr_i}{\sum_{j=1}^n sr_j} \quad (4)$$

$$lossrate = \sum_{i=1}^n wr_i \cdot l_i \quad (5)$$

Where wr_i is the weight of the total sales of vegetable products in the past two months, $l_1, l_2, \dots, l_n$ is the loss rate of vegetable products in the past two months, and $lossrate$ is the category loss rate in the past two months.

Analyze the relationship between the total sales of each vegetable category and the cost markup pricing. The unit cost of the vegetable category is the sum of all the unit costs from purchase to sale. This paper mainly considers the wholesale cost and the fixed costs involved in the process of selling vegetables (such as site fees, management fees, etc.). The calculation formula for unit cost is as follows:

$$perc = \frac{c + p \cdot order}{s_{total}} \quad (6)$$

Where $order$ is the order quantity, s_{total} is the total sales, c is the total fixed cost, and p is the wholesale price.

Through the derivation of Formula (7) and Formula (8), the replenishment formula can be obtained, as shown in Formula (9).

$$order = s_{total} + loss \quad (7)$$

$$loss = order \cdot lossrate \quad (8)$$

$$order = \frac{s_{total}}{1 - lossrate} \quad (9)$$

Calculate the cost-plus rate of the vegetable industry:

$$add = \frac{p_s - p}{p} \cdot 100\% \quad (10)$$

Where p_s is the sales price, and add is the cost-plus rate.

Finally, the cost-plus pricing formula for each vegetable category is established, as shown in Formula (11). Considering the loss rate, the daily replenishment formula is constructed, as shown in Formula (12).

$$price = perc \cdot (1 + add) = \frac{c + p \cdot \frac{S_{total}}{1 - lossrate}}{S_{total}} \cdot (1 + add) \quad (11)$$

$$order = \frac{S_{total}}{1 - lossrate} \quad (12)$$

2.2. ARIMA Model

Auto-Regressive Moving Average (ARIMA) Model [11] is a classical model in many times series analysis. Its basic idea is to regard the data sequence formed by the predicted object over time as a random sequence. The relationship between these random sequences reflects the scalability of the development of predictive objects. Its essence is to extract non-stationary deterministic information from time series by calculating differences [12]. Because vegetables are essential consumer goods, their sales are less affected by sales prices and can be regarded as consumer demand only related to time series. This paper uses the ARIMA Model to predict the wholesale price and sales volume of each vegetable category in the next week (July 1-7, 2023).

A time series is expressed as $\{X_t | t = 1, 2, 3, \dots, n\}$, and the time series value at time t is expressed as X_t . An ARIMA model is labeled as an ARIMA model (p, d, q) , where p is the number of the autoregressive term, d is the number of differences, and q is the number of moving averages [13]. In this paper, the ARIMA model is used to predict wholesale prices and sales, and the specific process is shown in Figure 1.

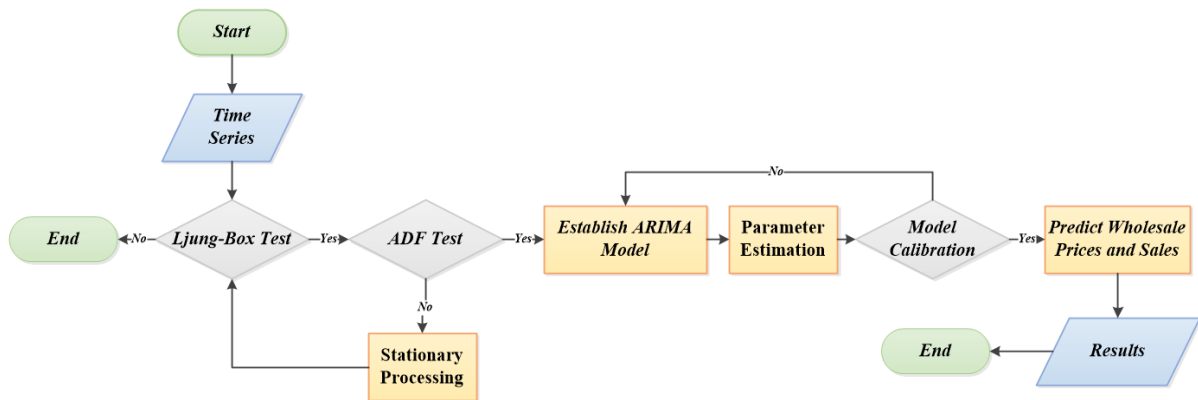


Figure 1. The flow chart of the ARIMA model

Step 1: The Ljung-Box Test is performed on the data. Only when the data is a non-white noise sequence, can the calculation be continued. The Augmented Dickey-Fuller Test (ADF) standard is used to determine whether the time series is stable. If it is stable, it enters Step2. If it is not stable, the data is calculated by difference, and the calculation order difference formula is as follows:

$$X_t - X_{t-k} = X_t - B^k X_t$$

$$BX_t = X_{t-1} \quad (13)$$

ADF test is performed on the new data again, and the process is repeated until the data is stable. The difference number is d when the data is stable.

Step 2: Calculate the autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF) and draw the image. The image of ACF attenuates and tends to 0 after q -order, and the image of PACF attenuates and tends to 0 after p -order. The parameters p and q are determined by the image.

Step 3: Establish a linear time series analysis model (AR):

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t \quad (14)$$

Where A is noise.

The moving average (MA) equation is established:

$$X_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (15)$$

The autoregressive average moving model (ARMA) is established:

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (16)$$

Step 4: Model Checking

The residual independent test and the normal test are used to judge the goodness of fit of the model. If both tests pass, it is proved that the model fits well. On the contrary, the fitting effect is poor. Using the accuracy of the test model, the calculation formula is as follows:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (17)$$

The wholesale price predicted by the ARIMA model is brought into the cost markup pricing Formula (11), and the sales price can be obtained. The daily sales volume predicted by the ARIMA model is brought into the daily replenishment Formula (12), and the daily replenishment volume can be obtained.

3. Results

In this paper, Matlab is used for programming. Firstly, the data set of single product wholesale price and sales volume is input, and the sum function is used to calculate the sum of sales volume. Then, calculate the wholesale price and sales volume of vegetable categories. Finally, the pricing model is calculated by regression method, and the pricing regression model of each category is shown in Table 1.

Table 1. Pricing regression model table of each category

Category	Regression Model
Cauliflowers	$price_A = 1.5106 \cdot p_A - 0.8755 \cdot \frac{1}{S_{total A}} + 0.0730$
Leafy Vegetables	$price_B = 1.6276 \cdot p_B + 5.1846 \cdot \frac{1}{S_{total B}} + 0.0618$
Peppers	$price_C = 1.3503 \cdot p_C - 3.6491 \cdot \frac{1}{S_{total C}} + 1.2193$
Solanaceae	$price_D = 1.5768 \cdot p_D - 0.0240 \cdot \frac{1}{S_{total D}} - 0.0389$
Edible Mushrooms	$price_E = 1.7358 \cdot p_E - 0.5943 \cdot \frac{1}{S_{total E}} - 0.8124$
Aquatic Rhizomes	$price_F = 1.5830 \cdot p_F + 0.4016 \cdot \frac{1}{S_{total F}} - 0.7782$

According to the calculation, the original observation data passed the Ljung-Box Test, and the time series model can be trained. However, it failed to pass the ADF test, so the first-order difference calculation was performed on it. Based on the ARIMA model, the adftest function is used to calculate and plot to determine the minimum difference order d when the time series is stable, and the arima _ test function is used to calculate the best p and q values to determine the ARIMA (p, d, q) model. The optimal parameters p, d and q of the wholesale price model and the sales model of each category are shown in Table 2 and Table 3.

Table 2. The optimal parameter table of the wholesale price model of each category

Category	p	d	q
Cauliflowers	3	1	3
Leafy Vegetables	3	1	5
Peppers	5	1	3
Solanaceae	4	1	1
Edible Mushrooms	4	0	5
Aquatic Rhizomes	5	0	5

Table 3. The optimal parameter table of each category sales model

Category	p	d	q
Cauliflowers	5	1	5
Leafy Vegetables	4	1	4
Peppers	5	1	5
Solanaceae	5	1	5
Edible Mushrooms	4	0	5
Aquatic Rhizomes	4	0	4

Taking the stationary processing of the wholesale price data of Cauliflowers as an example, it can be seen from Figure 2, Figure 3 and Figure 4 that the number of differences is 1 when the data is stationary, and the image oscillation of ACF attenuates and tends to 0 after the third order, and the image oscillation of PACF attenuates and tends to 0 after the third order.

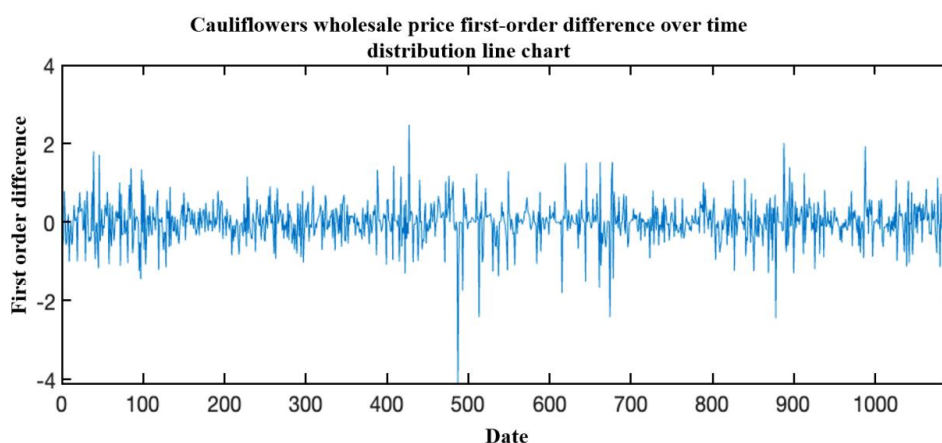


Figure 2. Cauliflowers wholesale price first-order difference over time distribution line chart

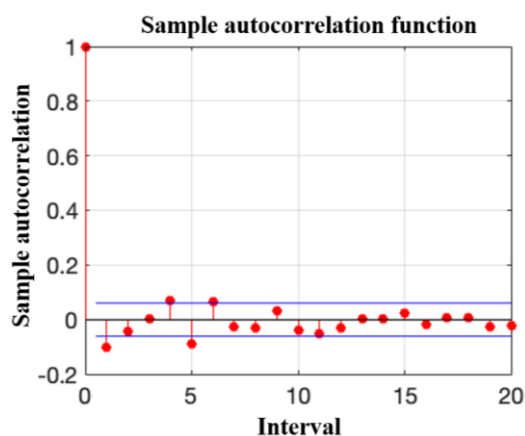


Figure 3. Autocorrelation function diagram

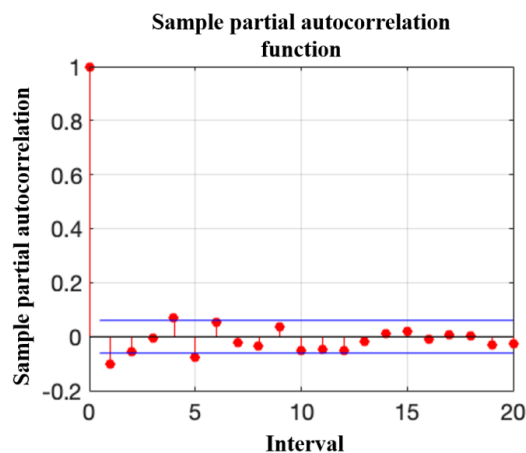


Figure 4. Partial autocorrelation function diagram

The residual independent test and normal test are carried out on the model. It can be seen from Figure 5 that the model passes two tests. In addition, the R^2 value of each category prediction model is greater than 0.7, indicating that the model fitting effect is good.

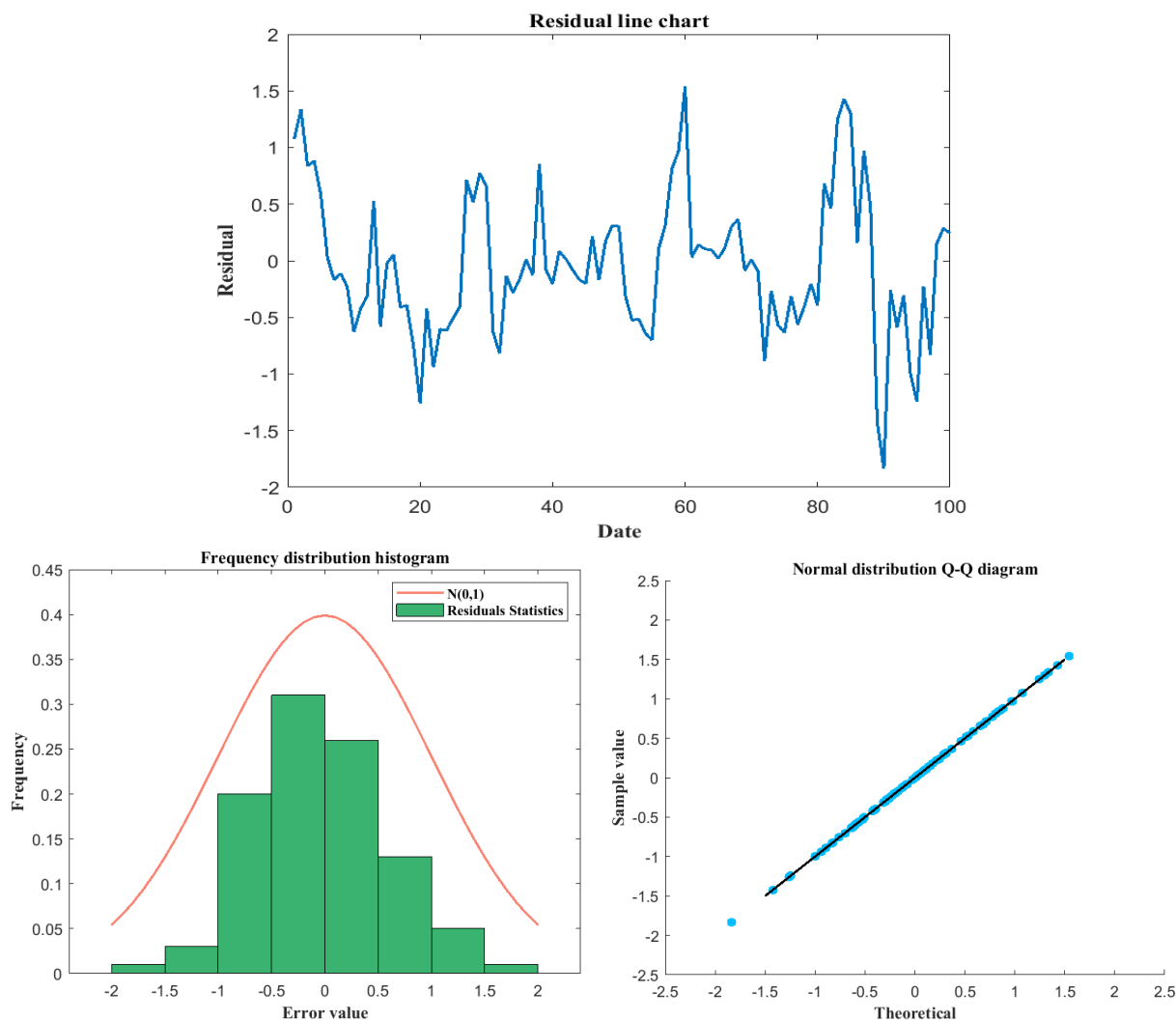


Figure 5. Residual independent test and normal test chart

Table 4. The R^2 value of each category prediction model

Category	Wholesale Price	Sales Volume
Cauliflowers	0.999878	0.999946
Leafy Vegetables	0.852807	0.992614
Peppers	0.999995	0.982713
Solanaceae	0.999306	0.950221
Edible Mushrooms	0.70316	0.776246
Aquatic Rhizomes	0.999912	0.823294

The determined optimal parameters are brought into the arima function, and the prediction unit step is set to 7. The trained model is used to predict the wholesale price and sales forecast value of the six categories. The results are shown in Figure 6 and Figure 7. According to the results, from July 1, 2023 to July 7, 2023, the daily wholesale price of Cauliflowers showed a downward trend, and the daily sales showed an upward trend. The daily wholesale price of Leafy Vegetables did not change much, remaining at about 3.25 yuan, and the daily sales showed a trend of decreasing first and then increasing. The daily wholesale price of Peppers showed an upward trend, and the daily sales showed a trend of decreasing first and then increasing. The daily wholesale price of Solanaceae showed an upward trend, and the daily sales showed a trend of decreasing first and then increasing. Edible Mushrooms' daily wholesale prices and daily sales fluctuations have no obvious rules. The daily wholesale price and daily sales of Aquatic Rhizomes showed a trend of rising first and then decreasing, and the daily wholesale price changed before the daily sales.

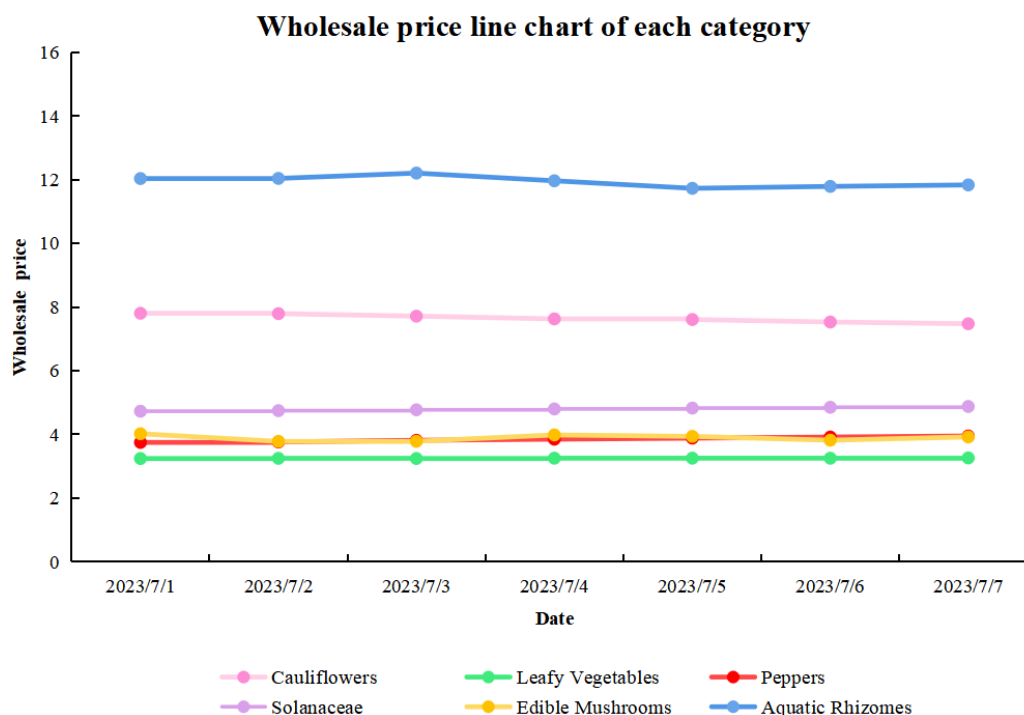


Figure 6. Wholesale price line chart of each category

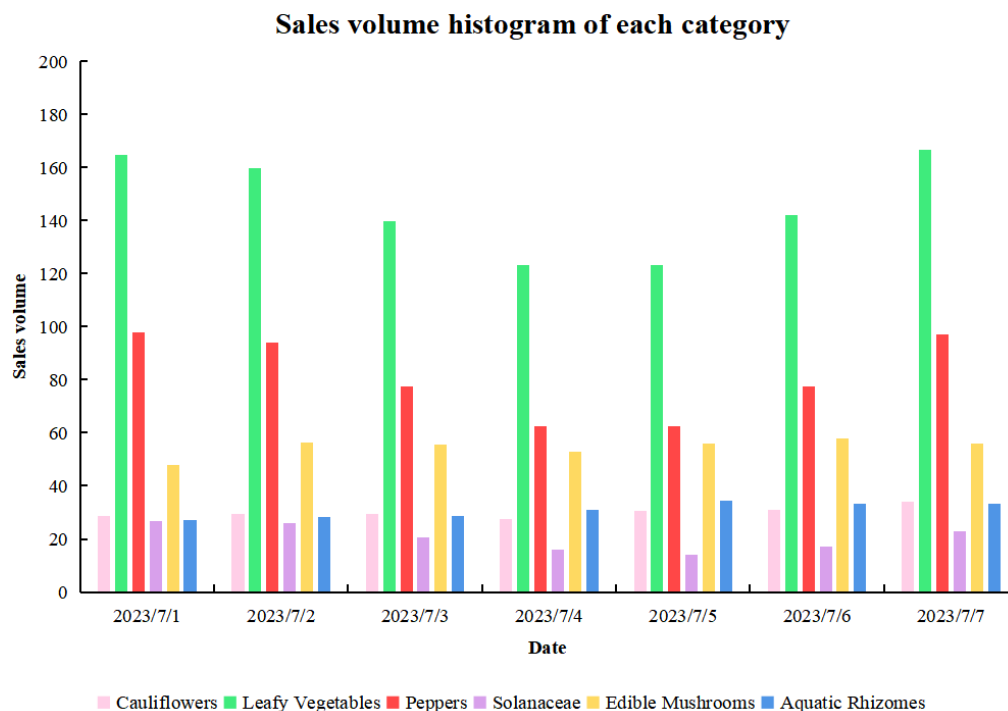


Figure 7. Sales volume histogram of each category

According to the formula, the loss rates of cauliflower, mosaic, pepper, eggplant, edible fungi and aquatic rhizome are 9.36 %, 10.20 %, 8.69 %, 6.19 %, 1.74 % and 14.75 %, respectively. The wholesale price, sales volume and loss rate are substituted into the pricing equation to obtain the daily replenishment quantity and pricing strategy of each category. The results are shown in Table 5 and Table 6.

Table 5. Daily replenishment forecast value table of each category

Date	Cauliflowers	Leafy Vegetables	Peppers	Solanaceae	Edible Mushrooms	Aquatic Rhizomes
2023/7/1	31.074	195.920	104.368	26.990	48.671	31.944
2023/7/2	26.050	184.697	106.158	27.416	57.351	33.258
2023/7/3	18.432	159.105	89.100	19.589	56.736	33.771
2023/7/4	13.010	124.702	72.992	15.261	53.867	36.546
2023/7/5	13.228	119.804	73.619	17.376	56.914	40.541
2023/7/6	19.014	143.406	83.454	19.714	58.930	39.126
2023/7/7	25.889	181.643	107.860	22.140	57.109	39.028

Table 6. Daily pricing forecast value table of each category

Date	Cauliflowers	Leafy Vegetables	Peppers	Solanaceae	Edible Mushrooms	Aquatic Rhizomes
2023/7/1	31.074	195.920	104.368	26.990	48.671	31.944
2023/7/2	26.050	184.697	106.158	27.416	57.351	33.258
2023/7/3	18.432	159.105	89.100	19.589	56.736	33.771
2023/7/4	13.010	124.702	72.992	15.261	53.867	36.546
2023/7/5	13.228	119.804	73.619	17.376	56.914	40.541
2023/7/6	19.014	143.406	83.454	19.714	58.930	39.126
2023/7/7	25.889	181.643	107.860	22.140	57.109	39.028

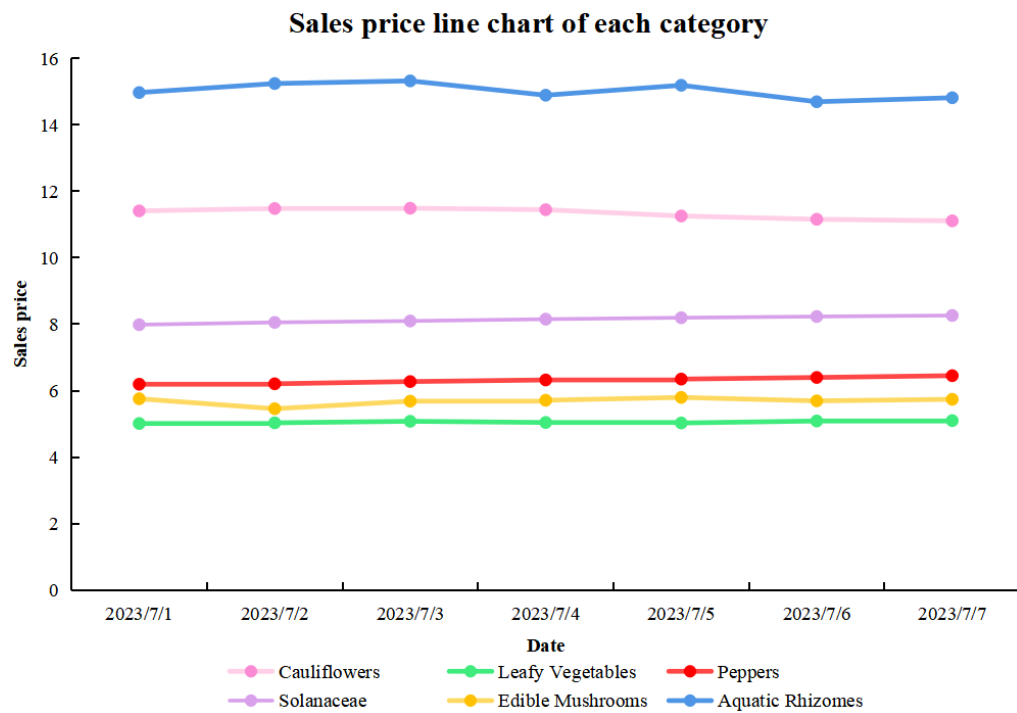


Figure 8. Sales price line chart of each category

Visualize the results. From Figure 8, it can be seen that the changes in sales pricing of each category are small, the fluctuation of sales pricing of Aquatic Rhizomes is large, and the fluctuation of Cauliflowers, Peppers and Solanaceae is small. According to Figure 9, the replenishment quantity of each category changes greatly. The replenishment quantity of Cauliflowers, Leafy Vegetables, Peppers and Solanaceae shows a trend of decreasing first and then increasing, and the replenishment quantity of Edible Mushrooms and Aquatic Rhizomes continues to rise.

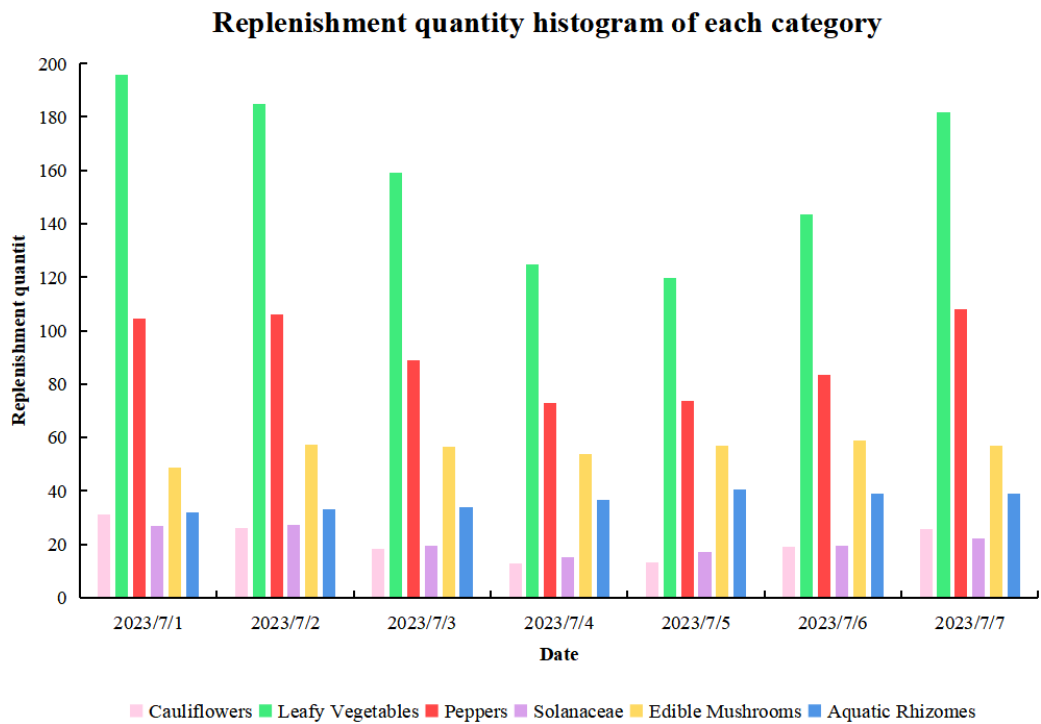


Figure 9. Replenishment quantity histogram of each category

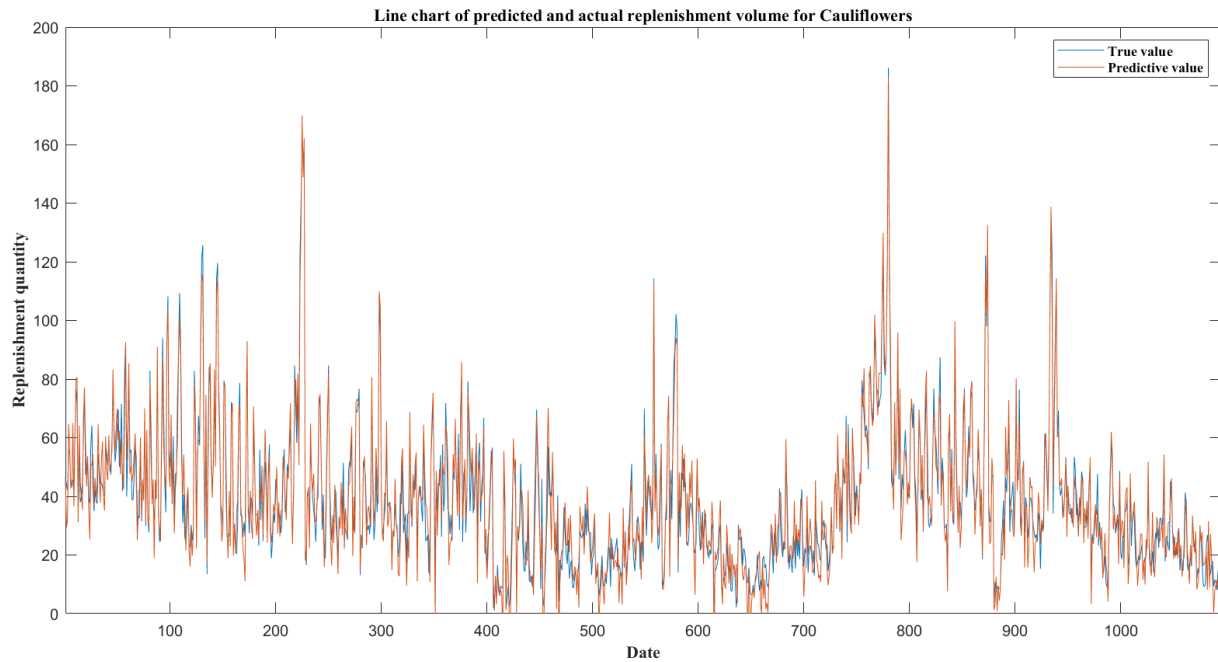


Figure 10. Line chart of predicted and actual replenishment volume for Cauliflowers

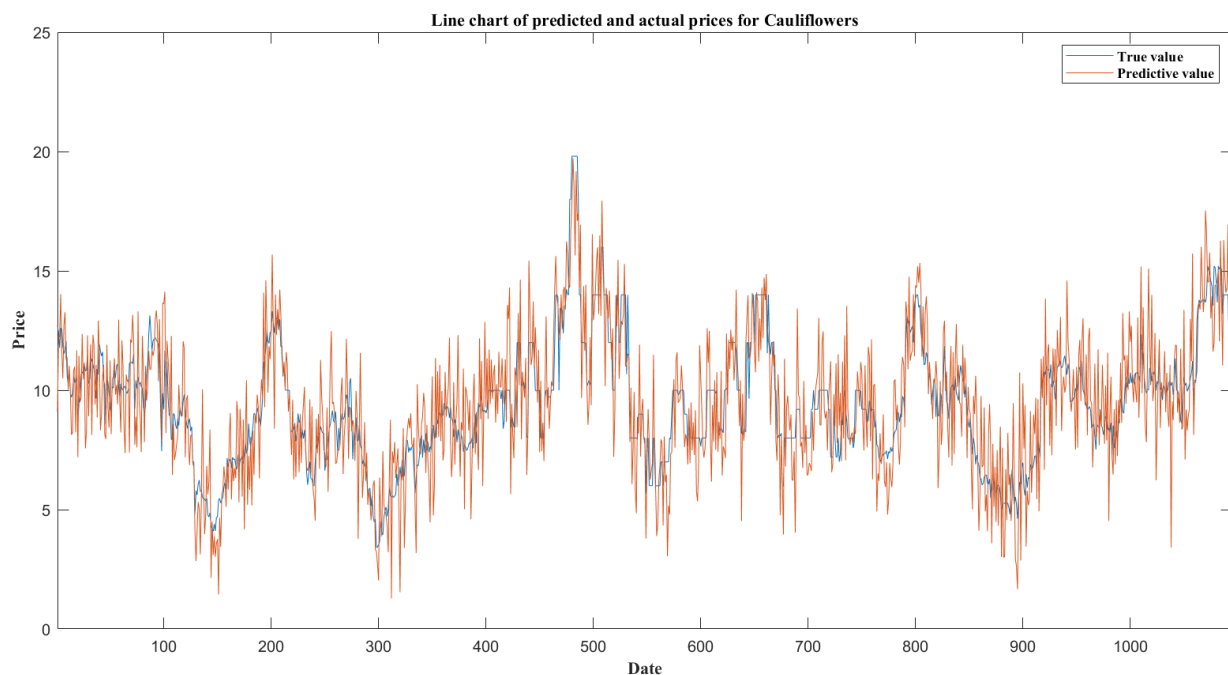


Figure 11. Line chart of predicted and actual prices for Cauliflowers

Taking Cauliflowers as an example, the sales price and replenishment quantity are predicted based on historical data, and the real value is compared with the predicted value. It can be seen from Figure 10 and Figure 11 that the predicted value is close to the real value, the change trend is more consistent, and there is no abnormal fluctuation. The calculation result is more reasonable, indicating that the model is more accurate and can be used for fresh supermarkets to formulate replenishment and pricing strategies.

4. Conclusion

In the replenishment and pricing of fresh supermarkets, it is particularly important to predict the future sales demand and sales unit price of vegetable categories. This paper uses the cost markup

pricing method to establish a pricing model, and uses the ARIMA model to predict the sales price and replenishment volume in the next seven days. The results show that the change of replenishment quantity of each category is large, but the fluctuation of sales pricing forecast is small. By comparing the prediction results of the model with the historical data of the same period, the predicted results are in line with the historical data and changes, which proves the rationality and prediction accuracy of the model proposed in this paper. The model constructed in this paper has passed the residual independent test and normal test, and R^2 is greater than 0.7, which has a good fitting effect and can provide a more accurate reference for subsequent vegetable pricing and replenishment decisions.

Acknowledgement

The authors gratefully acknowledge the data support from China Undergraduate Mathematical Contest in Modeling.

References

- [1] McLaughlin E W. The dynamics of fresh fruit and vegetable pricing in the supermarket channel [J]. Preventive Medicine, 2004, 39: 81 - 87.
- [2] Nahmias S. Perishable inventory systems [M]. Springer Science & Business Media, 2011.
- [3] Sainathan A. Pricing and replenishment of competing perishable product variants under dynamic demand substitution [J]. Production and Operations Management, 2013, 22 (5): 1157 - 1181.
- [4] Chen C Y, Lee W I, Kuo H M, et al. The study of a forecasting sales model for fresh food [J]. Expert Systems with Applications, 2010, 37 (12): 7696 - 7702.
- [5] Fan T, Xu C, Tao F. Dynamic pricing and replenishment policy for fresh produce [J]. Computers & Industrial Engineering, 2020, 139: 106127.
- [6] Bi Liyuan, Wu Sai, Chen Gang, Shou Li Dan, Chen Ke, Hu Tianlei. Database query cost prediction based on recurrent neural network[J]. Journal of Software, 2018, 29 (03): 799 - 810.
- [7] Dutta S, Maiti S. Price forecasting of agricultural products using arima models [J]. Indian Journal of Agricultural Marketing, 2021, 35 (2): 149 - 164.
- [8] Guilding C, Drury C, Tayles M. An empirical investigation of the importance of cost-plus pricing [J]. Managerial Auditing Journal, 2005, 20 (2): 125 - 137.
- [9] Barusman A R P, Yuliana T M, Mirfazli E. Analysis of implementation cost plus pricing method in the decision on the determination of product sales Prices [J]. Intern J Adv Sci Techn, 2020, 29 (06): 1832 - 1838.
- [10] Risvanti I. Cost plus pricing method in determining the selling price of the banana chips [J]. JOSAR (Journal of Students Academic Research), 2018, 3 (2): 50 - 60.
- [11] Nelson B K. Time series analysis using autoregressive integrated moving average (ARIMA) models [J]. Academic emergency medicine, 1998, 5 (7): 739 - 744.
- [12] Zhai M, Li W, Tie P, et al. Research on the predictive effect of a combined model of ARIMA and neural networks on human brucellosis in Shanxi Province, China: a time series predictive analysis [J]. BMC Infectious Diseases, 2021, 21 (1): 1 - 12.
- [13] Ariyo A A, Adewumi A O, Ayo C K. Stock price prediction using the ARIMA model [C]//2014 UKSim-AMSS 16th international conference on computer modelling and simulation. IEEE, 2014: 106 - 112.