Agricultural & Forestry Economics and Management (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/agrfem.2025.080107
ISSN 2616-2202 Vol. 8 Num. 1

Pricing and replenioring decision model of vegetable commodities based on historical data

Jiaqiang Xie, Pengzhan Niu, Chenglong Chao, Huanzheng Zhu

School of Mechanical and Electronic Engineering, Shandong Jianzhu University, Jinan, 250101, China

Keywords: Prediction Model, ARIMA Time Series Analysis, Linear Programming

Abstract: In modern retail, vegetables, as daily consumer goods, have strong seasonality and demand fluctuations, and at the same time have a short shelf life and are easy to wear out. Retailers need to optimize replenishment and pricing strategies to ensure adequate supply and control inventory depletion. How to maximize the profit of the supermarket on this basis has become an important problem that the retail industry needs to solve. We analyzed the relationship between the total sales volume and cost-plus pricing of each vegetable category, established the ARIMA model and the revenue maximization model, and gave the daily replying volume and pricing strategy of each vegetable category in the next week (July 1-7, 2023). According to the available varieties in the past week, under the condition that the order quantity of each item is in the range of 27 to 33, and the order quantity of each item is more than 2.5 kg, the linear programming is used to optimize the decision under the constraints to maximize the total revenue, and the replying quantity and pricing strategy of each item on July 1 are given. By optimizing pricing and replying strategies, retailers can better respond to market changes, improve their market competitiveness and profitability, and promote the development of intelligent retail industry, which has important academic significance and practical value.

1. Introduction

Through accurate replenishment decisions, retailers and wholesalers can more effectively connect with farmers or producers, optimize the production and supply chain planning of vegetables. This can drive the digital transformation in the retail and agricultural sectors and achieve intelligent management, promoting the modernization of the agricultural industry. If the predicted replenishment volume and pricing for vegetable varieties have too large an error, it may have negative impacts on multiple aspects, especially in the areas of supply chain, sales, and profits. It can lead to losses in sales revenue and profits, increased inventory management costs, and increased supply chain volatility, potentially triggering larger-scale logistics and supply problems.

Wang K, Su K, Li H applied the Random Forest (RF) model to the prediction of pricing and replenishment decisions for vegetable products based on historical data [1]. Liang Y, Li Y, Chen X and Jiang Y, Li X respectively utilized Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory Recurrent Neural Network for automatic pricing and prediction of vegetable products [2-3]. Lu Z, Wang Y, Liu K employed the DPCCA model in the prediction of pricing and replenishment decisions for vegetable products based on historical data [4].

Previous replenishment forecasting and pricing strategies are mainly based on LSTM, Random Forest and DPCCA models. LSTM model requires a lot of memory and computing power due to its deep structure and cyclic calculation. Moreover, it needs to be iterated step by step, the training time is long, and it is easy to overfit when the data amount is small or the noise is large. Each decision tree in random forest is only based on local information, ignoring the complex relationship between variables. The training of a large number of decision trees requires high computing resources, and despite the integration methods, it is still easy to overfit in noisy data. The DPCCA model assumes that the potential class is fixed, ignores market dynamics, and requires high computational resources for large-scale matrix operations and optimization processes. This paper uses the ARIMA model for time series analysis and linear programming to optimize decision variables, and finally provides specific replenishment quantities and pricing strategies. The ARIMA model is suitable for non-stationary time series, has good short-term prediction performance and certain flexibility, and can handle various types of time series data. The linear programming model has a clear and concise mathematical structure, wide application fields, mature and efficient solution methods, sensitivity analysis capabilities, and the ability to handle multiple constraints and objectives. This is helpful for decision-making and optimization under complex conditions.

Data source: https://www.mcm.edu.cn/. The data mainly include detailed sales data and wholesale prices of vegetable commodities.

2. Establishment of model

2.1 The process of establishing the ARIMA prediction model

2.1.1 Descriptive statistical analysis

(1) Calculate the sales volume of each category

By calculating the total sales volume of each category, we can understand which categories have good sales performance, and help supermarkets determine which categories are the focus of sales, so as to optimize inventory management. The formula for total sales is as follows:

$$TotalSales_i = \sum_{t=1}^{T} S_{i,t} \tag{1}$$

where $S_{i,t}$ is the sales volume of the category in time, and T is the total number of days in the time period.

By calculating the total amount of sales over a period of time, it is possible to identify which categories are the main selling varieties and which categories need to be optimized.

(2) Calculate the average selling price and average cost for each category

By calculating the average selling price and average cost, we can understand the profitability of each category and provide a basis for pricing strategy.

Average selling price:

$$P_i = \frac{\sum_{t=1}^{T} P_{i,t} \times S_{i,t}}{\sum_{t=1}^{T} S_{i,t}} \tag{2}$$

where $P_{i,t}$ is the selling price of the category in time

Average cost:

$$C_i = \frac{\sum_{t=1}^{T} C_{i,t} \times S_{i,t}}{\sum_{t=1}^{T} S_{i,t}} \tag{3}$$

where $C_{i,t}$ is the cost of the category in time (that is, the wholesale price).

(3)Analyze the cost markup

Cost markup analysis can help supermarkets evaluate whether the pricing strategy of each category is reasonable, so as to ensure reasonable profit on the basis of cost. The formula is as follows:

$$m_i = \frac{P_i - C_i}{C_i} \tag{4}$$

where m_i is the markup, P_i represents selling price, and C_i means cost.

The markup reflects the percentage of profit added to the cost basis of the supermarket, which is the key indicator to formulate the pricing strategy. By calculating the average selling price, average cost and markup of each category, it is possible to identify which categories are priced with high profitability and which categories may need to adjust their pricing to improve profitability.

2.1.2 Demand forecasting

(1) Formula of ARIMA model

To accurately predict the daily sales volume of each vegetable category for the next week (July 1 to 7, 2023), ARIMA model can be used for time series analysis. The ARIMA model combines three components: autoregressive (AR), difference (I) and moving average (MA).

The specific formula of ARIMA model is as follows:

$$Y_t = c + \sum_{i=1}^{p} \emptyset_i Y_{t-i} + \sum_{j=1}^{q} \theta_j Y_{t-j} + \varepsilon_t \tag{5}$$

where Y_t is the predicted value of t, c is the constant term (intercept), $\emptyset_i$ is the autoregressive coefficient (AR part), θ_j is the moving average coefficient (MA part), and ε_t is the error term.

Model parameter selection:

·**p (autoregressive order) :** determined from the partial autocorrelation function (PACF) plot.

·**d (number of differences) :** determined by unit root test (such as ADF test).

·**q (moving average order) :** determined from the autocorrelation function (ACF) plot.

(2) Model fitting and prediction

① Data preparation and stationarity detection

a) Check the stationarity of time series

● Stationarity: A time series is stationary, meaning that its statistical properties (such as mean, variance, autocorrelation) do not change over time. Many time series are non-stationary in their original form and need to be made stationary by differencing.

● ADF test: This is a common statistical test used to test whether a series is stationary. If the p-value of the test result is small (usually less than 0.05), it indicates that the time series is stationary.

b) Differential operation

If the time series data has a trend, it needs to be transformed into a stationary series by difference operation. The first difference formula is as follows:

$$Y_t' = Y_t - Y_{t-1} \tag{6}$$

Number of differencing: The number of differencing refers to how many differencing operations are needed to make the series stationary.

② Determination of autoregressive order p

The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were used to determine the model parameters p and q.

a) Autocorrelation function (ACF)

● The autocorrelation function is used to measure the correlation between the current value and the value of previous periods.

● If the series is stationary after differencing, the order q of the moving average part can be determined by looking at the ACF plot.

b) Partial autocorrelation function (PACF)

③ Determination of moving average order q

a. Autocorrelation function (ACF)

● The ACF plot is not only used for the determination of d, but can also help determine the order q of the moving average part.

● Determine the value of q: Observe the ACF plot and find the position of the truncation after the first significant autocorrelation coefficient; this lag order is the value of q.

④ Fitting the ARIMA model

The determined parameters p, d, and q are used to fit the ARIMA model, and the model is trained to predict future sales volume.

⑤ Forecast future sales

Using the trained ARIMA model, forecast the sales volume S_i for the next 7 days.

2.1.3 The loss rate is included in the calculation of replenishment volume

Accurately calculating the actual replenishment quantity Q_i for each category is the key to ensure adequate inventory when taking wastage into account. The calculation method is as follows: based on the predicted sales volume S_i, combined with the loss rateθ_i, the actual quantity Q_i to be replanted is calculated:

$$Q_i = \frac{S_i}{1-\theta_i} \tag{7}$$

This formula ensures that the supermarket can still have enough inventory to meet the expected demand after taking wastage into account, avoiding out-of-stock situations caused by wastage.

2.1.4 Cost-plus pricing strategy

To influence sales volume and achieve revenue maximization, adjusting the pricing strategy is crucial. The pricing method is as follows: the cost-plus pricing method, where the selling price P_i is calculated based on the wholesale cost C_i and the set markup rate m_i.

$$P_i = C_i \times (1 + m_i) \tag{8}$$

The selling price P_i is adjusted to optimize the sales volume S_i based on the price elasticity of market demand.

2.1.5 Revenue maximization model

In order to maximize the total revenue of the supermarket, it is necessary to fully consider the pricing strategy and replenishment plan, such as pricing, replenishment volume and loss rate.

$$R = \sum_{i=1}^{n} [(P_i - C_i) \times S_i] \tag{9}$$

When calculating revenue, R is the total revenue S_i of the supermarket, the actual sales volume of category i,P_i is the selling price of category i, and C_i is the cost of category i(i.e. the wholesale price).

By dynamically adjusting pricing and replenishment strategies, Shangsuper can not only maximize current revenue, but also improve long-term market competitiveness and profitability.

2.2 The process of establishing profit maximization model of linear programming

2.2.1 Data preparation and preliminary analysis

(1) Data extraction

First, the relevant information is extracted from the sales data from June 24 to June 30, 2023, including the sales volume, selling price, wear and tear rate of each item.This data will serve as the basis for predictions and decisions. Among them, the sales volume data is used to analyze the market demand trend of each product, the sales price data is used to formulate the subsequent pricing strategy, and the depletion rate data is used to help calculate the actual needed replenishment volume to avoid waste.

(2) Data cleaning and screening

Clean the data, deal with missing values and outliers, and ensure the accuracy and integrity of the data. At the same time, the items that were actually sold during the period from June 24 to 30 will be selected as candidates for possible sales on July 1.

2.2.2 Item selection and replenishment plan

(1) Single item selection criteria

Due to limited sales space, the number of individual items must be controlled between 27 and 33. We can filter through several criteria:

● Sales performance: Priority is given to items with large and stable sales volume, which have a high demand in the market.

● Profit margin: When choosing a single product, it is also necessary to consider the profit margin, and preferentially choose a single product with a higher markup rate to improve the overall profitability.

● Demand elasticity: Considering the price elasticity of market demand, choose a single product with stable demand when the price changes to ensure the stability of income.

(2) Calculation of replenishment quantity

After determining the item mix, it is necessary to calculate a reasonable replenishment amount for each item. The minimum display size requirement is that the ordered quantity of each item must meet the minimum display size requirement of 2.5 kg, which can be calculated by the following formula:

$$Q_i = max\left(2.5, \frac{\widehat{Q}_i}{1 - \theta_i}\right) \tag{10}$$

where Q_i is the replenishment volume of item i,$\widehat{Q}_i$ is the predicted sales volume of item i, and θ_i is the loss rate of item i.

Through this formula, it is ensured that each item can still meet the market demand after loss, and at the same time meet the minimum display quantity requirements.

2.2.3 Formulation of pricing strategy

(1) Cost plus pricing

Using the cost plus pricing method, the base selling price P_i is calculated based on the wholesale cost C_i of the item and the expected profit margin m_i:

$$P_i = C_i \times (1 + m_i) \tag{11}$$

where P_i is the base selling price of item i, C_i is the wholesale cost of item i, and m_i is the expected profit margin of item i.

(2) Price elasticity adjustment

In order to ensure reasonable pricing and market competitiveness, it is also necessary to consider the price elasticity of market demand and the pricing strategies of competitors. After determining the basic price, it is also necessary to consider the price elasticity of market demand, and adjust the price according to the actual market conditions to optimize the sales volume and profit mix.

① Price elasticity analysis

Price elasticity reflects the impact of price changes on the demand. By analyzing the historical sales data, the price elasticity coefficient E_i of each item can be estimated.

② Price elasticity formula

$$E_i = \frac{\Delta D_i / D_i}{\Delta P_i / P_i} \tag{12}$$

where E_i is the price elasticity coefficient of item i; ΔD_i is the change in quantity demanded of item i; D_i is the initial quantity demanded of item i; ΔP_i is the change in price of item i; P_i is the initial price of item i.

(3) Dynamic pricing strategy

Dynamic pricing is a strategy that adjusts prices based on real-time market information. By analyzing historical data and current market dynamics, supermarket can continuously optimize prices to adapt to changing market demand and maximize sales and profits [5].

Based on historical sales, inventory levels and attrition rates in existing data, we can make simplified dynamic pricing adjustments:

$$P_i' = P_i \times \big(1 + f(S_i, I_i, \Delta D_i)\big) \tag{13}$$

where: P_i' is the adjusted selling price of item i, P_i is the base selling price of item i, and $f(S_i, I_i, \Delta D_i)$ is the price adjustment function, reflecting the influence of the quantity demanded, inventory level and rate of change of demand.

$$f(S_i, I_i, \Delta D_i) = \alpha \frac{S_i}{\overline{S_i}} + \beta \frac{\overline{I_i}}{I_i} + \gamma \frac{\Delta D_i}{\overline{\Delta D_i}} \tag{14}$$

where S_i is the historical average demand, Ii is the historical average inventory level, ΔD_i is the historical average demand change rate, α,β , γ are the weight coefficients, determined by regression analysis.

Through the comprehensive application of cost plus pricing, price elasticity adjustment and dynamic pricing strategy, supermarket can flexibly adjust prices under different market environments to achieve the best combination of sales volume and profit. Continuous monitoring and feedback adjustments ensure the effectiveness of pricing strategies, enabling supermarkets to respond to changing market demands and maintain an edge over the competition.

2.2.4 Optimization algorithm: Use linear programming to maximize revenue

In order to maximize the total revenue of the supermarket, we need to comprehensively consider pricing, replenishment volume and loss rate when making pricing strategy and replenishment plan. By using linear programming, we can optimize these decision variables under given constraints, thus maximizing the benefits.

We need to define the decision variables, objective functions, and constraints, assuming we have m candidate items, through the following steps.

(1) Definition of decision variables

x_i: Whether to select item i($x_i = 1$ indicates yes,$x_i = 0$ indicates no).

Q_i: Replenishment quantity of item i.

(2) Model expression

① Objective function

The goal is to maximize the total revenue on July 1 by optimizing pricing and replenishment strategies. It can be expressed as:

$$MaximizeR = \sum_{i=1}^{m} x_i \times (P_i - C_i) \times Q_i \tag{15}$$

where R is the total revenue, P_i is the selling price of item i, C_i is the cost of item i, and m is the final number of items selected.

② Constraint conditions

In the process of solving, the following constraints must be considered:

a) Single product selection constraints:

$$\textstyle\sum_{i=1}^{m} x_i \geq 27 \quad and \quad \sum_{i=1}^{m} x_i \leq 33 \tag{16}$$

b) Minimum display quantity constraint:

$$Q_i \geq 2.5 \times x_i \quad for\ all\ i \tag{17}$$

c) Sales space constraints:

$$\sum_{i=1}^{m} Q_i \leq S_{max} \tag{18}$$

d) Budgetary constraints (optional) :

$$\sum_{i=1}^{m} C_i \times Q_i \leq B_{max} \tag{19}$$

3. Results

3.1 Prediction results of ARIMA model

By establishing ARIMA model and revenue maximization model, the daily replenishment amount and pricing strategy of each vegetable category in the coming week (July 1-7, 2023) are given. The output results of the model take aquatic rhizomes and solanum as examples, Figure 1 shows the average pricing strategy for each category.

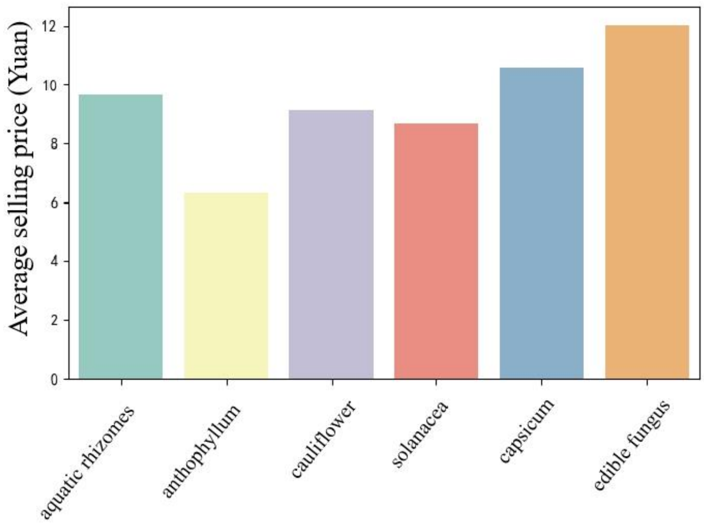
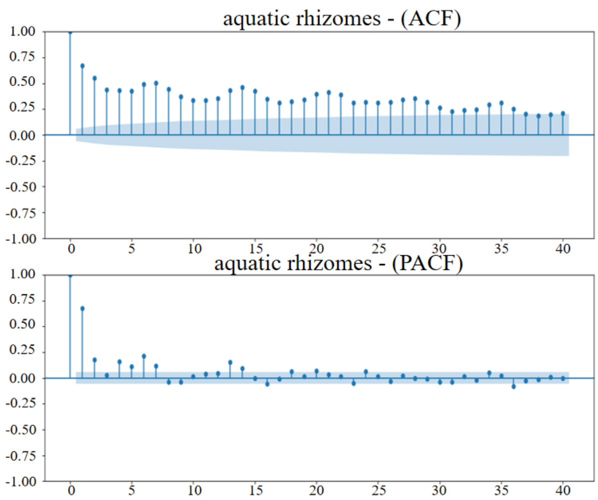


Figure 1. Average selling price histogram by category



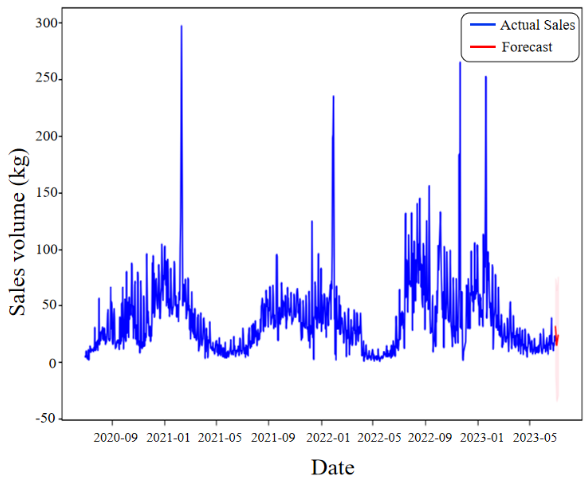
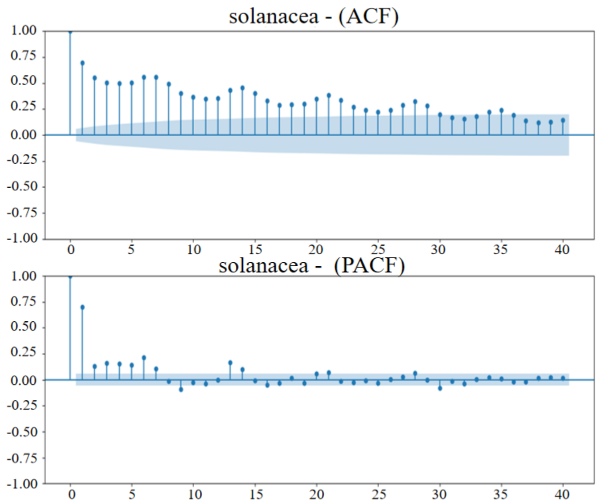


Figure 2. Aquatic rhizomes

Figure 3. Sales forecast chart for next seven days



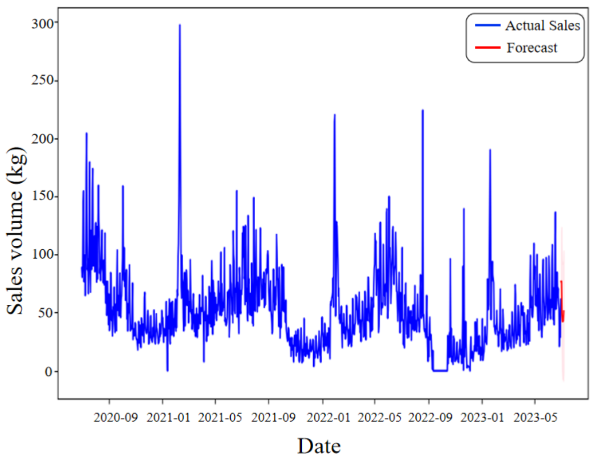


Figure 4. Solanacea

Figure 5. Sales forecast chart for next seven days

Three parameters that fit the ARIMA model can be determined from Figures 2 and 3, and the results predicted by the model are shown in Figures 4 and 5. Specific results were obtained after model prediction. Aquatic rhizomes were taken as an example, as shown in Table.1.

Table.1. Forecast sales table of aquatic rhizomes

Aquatic rhizomes	Daily forecast sales (kg)	Solanacea	Daily forecast sales (kg)
2023-07-01	30.98	2023-07-01	30.69
2023-07-02	26.16	2023-07-02	30.76
2023-07-03	16.055	2023-07-03	19.39
2023-07-04	15.06	2023-07-04	17.00
2023-07-05	18.16	2023-07-05	17.30
2023-07-06	20.53	2023-07-06	17.14
2023-07-07	23.34	2023-07-07	20.47

As can be seen from other output results, the total forecast sales volume of aquatic roots is 150.3kg, and the forecast sales volume is on a downward trend from July 1 to July 4. The highest sales volume was 30.98 on July 1, while the lowest was 15.06kg on July 4. From July 4 to July 7, the forecast sales are on an upward trend. Sales volume on July 7 was 23.34kg, a significant increase from July 4.

The total forecast sales volume of nightshade is 152.7kg, and from July 1 to July 4, the forecast sales volume of Nightshade shows a downward trend. From July 4 to July 7, the forecast sales are on an upward trend. Sales on July 1 and July 2 were relatively high and close, at 30.69kg and 30.76kg respectively. On July 3, sales fell sharply to 19.39kg. Sales continued to decline to 17.00kg on July 4. Sales on July 5 and July 6 were relatively stable at 17.30kg and 17.14kg respectively. On July 7, sales increased significantly to 20.47kg.

This can be used as a reference for production and inventory management to ensure that there is sufficient product supply during the week, while avoiding excessive inventory. Based on these forecast data and analysis, enterprises can rationally arrange production, inventory and marketing activities to meet market demand and optimize sales strategies.

3.2 Prediction results of linear programming benefit maximization model

According to the available varieties from June 24 to 30, 2023, under the condition that the number of individual items is in the range of 27 to 33, and the ordered quantity of each item meets the minimum display quantity of 2.5 kg, the linear programming benefit maximization model prediction is made, and the single item replenishment quantity and pricing strategy are obtained on July 1.Some of the output results are shown in Table 2.

Under the constraints of the number of available items for sale and the order quantity of each item, a linear programming model for maximizing profits was used to obtain the profit margin, wholesale price, comprehensive score, and predicted replenishment quantity. These data can help merchants better understand the profit potential and market demand of different agricultural products, thereby making reasonable purchasing and sales decisions. For example, leafy vegetables have a relatively high profit margin and may be a more profitable choice, while aquatic root and stem vegetables and edible fungi have relatively low profit margins, and other factors such as market demand and replenishment quantity need to be comprehensively considered to determine whether to purchase.

Table.2. Forecast replenishment scale

Item name	systematic name	profit margin	Wholesale price (Yuan/kg)	comprehensive evaluation	Forecast replenishment (kg)
boletus aereus	edible mushrooms	0.38	65.98	0.491	0.346
Sichuan red toon	anthophyllum	0.73	32.54	0.476	0.273
Honghu lotus root belt	aquatic rhizomes	0.26	29.43	0.453	0.362
Artemisia stelleriana	anthophyllum	0.27	25.96	0.445	0.300
Luffa tip	anthophyllum	0.62	14.98	0.444	0.485
foldleaf daylily root	anthophyllum	0.94	5.00	0.438	0.189
Tricholoma matsutake	edible mushrooms	0.58	40.67	0.438	0.430
Colorful Pepper (2)	capsicum	0.65	9.14	0.435	0.255
water caltrop	aquatic rhizomes	0.65	8.18	0.435	0.440
Wild pink lotus root	aquatic rhizomes	0.62	9.76	0.435	0.260
Small green vegetables (portion)	anthophyllum	0.83	0.59	0.435	1.124
Tall Melon (1)	aquatic rhizomes	0.82	5.65	0.434	0.424
Cabbage heart (portion)	anthophyllum	0.81	1.15	0.434	1.124
Green Eggplant (1)	solanacea	0.88	1.63	0.434	0.774

4. Conclusions

Based on the historical data of various vegetable categories, this paper established an ARIMA model and a revenue maximization model to determine the daily replenishment volume and pricing strategy for each vegetable category in the coming week, aiming to maximize the revenue of the supermarket. The results were presented. Under the condition that the total number of available items is between 27 and 33 and the order quantity of each item meets the minimum display quantity of 2.5 kilograms, a linear programming model was established. Under the premise of meeting the market demand for various vegetable products, the replenishment quantity and pricing of each item on July 1st were given based on the available varieties from June 24th to 30th, 2023. Finally, the results obtained from the model were analyzed. By establishing a mathematical model to optimize the replenishment and pricing strategies of vegetables in supermarkets, it is possible to accurately predict demand fluctuations, avoid overstocking or understocking, reduce waste and maximize revenue. Reasonable replenishment and pricing not only improve operational efficiency and reduce costs, but also help cope with seasonal changes, weather fluctuations and other market fluctuations, optimize supply chain management and reduce logistics costs. Overall, scientific prediction and decision-making can enhance profitability, customer satisfaction and market competitiveness, promote the sustainable development of supermarkets and have multiple practical significances.

The pure mathematical formula reasoning makes the model complex in solving and it is difficult to obtain the optimal solution in a short time. The derivation process is overly idealized. In practical applications, the return rate may also be an important factor, but this paper failed to consider the influence of these factors, which to some extent affects the accuracy of the model. This model has good adaptability and optimality in the "Crossing the Desert" game and can be extended to other related problems that require dynamic programming to achieve maximum revenue.

References

[1] Jiang Y, Li X. Automatic Pricing and Replenishment Decision-Making for Vegetable Commodities Based on Bi-directional Long Short-Term Memory Recurrent Neural Networks and Markov Prediction Models[J]. Academic Journal of Science and Technology, 2023, 7(3): 69-73.

[2] Liang Y, Li Y, Chen X. Prediction and Replenishment Decision Making for Automatic Pricing of Vegetable Commodities Based on LSTM Models[J]. Academic Journal of Science and Technology, 2023, 8(1): 264-268.

[3] Wang K, Su K, Li H. Random Forest-Based Restocking and Pricing Prediction for Vegetable Items[J]. Frontiers in Business, Economics and Management, 2023, 11(2): 352-356.

[4] Lu Z, Wang Y, Liu K. Research on Pricing and Replenishment Strategies for Supermarkets Based on Revenue Maximization [J]. Journal of Education, Humanities and Social Sciences, 2024, 37: 148-157.

[5] Shin D, Vaccari S, Zeevi A. Dynamic pricing with online reviews [J]. Management Science, 2023, 69(2): 824-845.