# Research On Pricing Strategies Based on Time Series Analysis Model and Greedy Algorithm Model

Hongfeng Li [1, *], Yuyan Jia [2], Yuzhu Zhang [3]

[1] School of Information, Shanxi University of Finance and Economics, Shanxi, China, 030006

[2] School of statistics, Shanxi University of Finance and Economics, Shanxi, China, 030006

[3] School of Economics and Finance, Shanxi University of Finance and Economics, Shanxi, China, 030006

* Corresponding Author Email: 18535828749@163.com

**Abstract.** The issue of replenishment and pricing arrangements for vegetable categories and individual items in future supermarkets is crucial. This article is based on a time series analysis model to analyze the correlation between different categories of single products, and constructs a dual objective greedy algorithm optimization model, constrained by maximizing profits and minimizing operational loss rates. Predict the total replenishment amount and pricing strategy for 50 types of vegetables, including categories such as cauliflower, leaves, chili peppers, eggplants, edible mushrooms, and underwater acoustic root canals. Next week's predicted sales and prices are: cauliflower 0.381 kg, 11.7331 yuan/kg; Leaves 5.9-6.084 kg, 9.031-8.504 yuan/kg; Chili peppers range from 8.327 to 5.8 kilograms, with a price of 15.632 to 14.793 yuan/kilogram; Eggplant 10.962-11.321 kg, 7.843-7.861 yuan/kg; Edible mushrooms 5.5118-5.374 kg, 12.323-11.476 yuan/kg; The underwater sound root canal costs 3.321-3.378 kilograms, 25.619-15.981 yuan/kilogram.

**Keywords:** Time series analysis model, Optimization model of dual objective greedy algorithm, Big Data.

## 1. Introduction

Fresh fruits and vegetables are a rich source of micronutrients. [1]. Consumers are more and more inclined to buy high-quality products [2]. In fresh supermarkets, the shelf life of vegetable products is usually short, and their quality gradually decreases over time, which has a negative impact on sales. If most varieties are not sold within one day, they will no longer be suitable for sale the next day. Therefore, supermarkets usually develop daily replenishment plans based on historical sales data and demand for each vegetable product. Due to the large variety of vegetables sold in supermarkets and varying origins, and the purchase transaction time for vegetables is usually from 3:00 a.m. to 4:00 a.m., merchants are uncertain about specific individual items and purchase prices when placing orders, making it difficult to make replenishment decisions for each vegetable category on that day. So, this paper hopes to use mathematical methods to analyze existing historical data, establish mathematical models, and help supermarkets better arrange automatic pricing of goods and make reasonable replenishment decisions.

The automatic pricing and replenishment decision-making of products has always been a hot issue in vegetable procurement. In recent years, extensive research has been conducted on the joint optimization of commodity pricing and replenishment from different perspectives. Neda et al. [3] and Mahmoodi. [4] established a replenishment pricing model with piecewise time-varying quantity consumption rate and sensitive demand to price and inventory. Tijun Fan et al. [5] established a dynamic pricing and replenishment policy considering freshness and order quantity. Yongrui Duan et al. [6] Established a replenishment pricing model that considers both quantity and quality losses. They all adopt direct methods to solve retailers' replenishment pricing strategies to maximize profits.

In summary, existing research on maximizing profits through direct methods is relatively mature, but there are relatively few studies that consider the combination of profit and operational loss rate. When expressing the demand function, it is mostly limited to describing the internal relationship between quality loss, product quality, and demand. There is little research on further considering total sales and cost-plus pricing based on these three factors. However, most existing research is based on static models, which assume that only one or more supply cycles have the same supply volume. This text studied the wholesale prices and sales flow details of the past three years, and the dynamic programming method is more scientific.

## 2. Forecasting based on time series data

### 2.1. Time Series Modeling Principle of time series model

Time series modeling and forecasting have fundamental importance to various practical domains. [7] In particular, the ARIMA model has demonstrated its outperformance in precision and accuracy of predicting the next lags of time series.[8] Time series refers to a numerical sequence that arranges the indicator values of a certain phenomenon in chronological order. Time series analysis can be roughly divided into three parts: describing the past, analyzing patterns, and predicting the future. In this question, we mainly use the ARIMA model of time series analysis. Autoregressive model (AR (p)).

Describe the relationship between the current value and the historical value and use the historical data of the variable itself to predict it. It must meet the requirements of stationarity and is only applicable to predicting phenomena related to the previous period of the variable (autocorrelation of time series).

### 2.1.1 Principle of time series model

Moving Average Model (MA (q)): The moving average model focuses on the formula definition of the cumulative q-order autoregressive process of the error term in the autoregressive model:

$$y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{1}$$

The current value of the time series is not related to the historical value but only relies on the linear combination of historical white noise. The moving average method can effectively eliminate random fluctuations in prediction.

Autoregressive Moving Average Model (ARMA (p, q))

The combination of autoregressive and moving average is defined by the formula:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{2}$$

This equation indicates that if the sequence is stationary, that is, its behavior will not change over time, then we can predict the future based on the past behavior of the sequence.

Differential autoregressive moving average model ARIMA (p, d, q)

By combining the autoregressive model (AR), moving average model (MA), and differential method, we obtain the differential autoregressive moving average model ARIMA (p, d, q), where p is the autoregressive term, q is the number of moving average terms, and d is the number of differences made when the time series becomes stationary.

### 2.1.2 Modeling steps of ARIMA model

1. Conduct stationarity testing on sequence mapping and observe whether the sequence is stationary; For non-stationary time series, first perform d-order difference and convert it into a stationary time series.

2. After the first step of processing, a stationary time series has been obtained. To obtain the autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF) of a stationary time series, analyze the autocorrelation graph and partial autocorrelation graph to obtain the optimal order p, q:

3. Based on the d, q, and p obtained above, obtain the ARIMA model. Then start model validation on the obtained model.

Firstly, we need to conduct a stationarity test.

### 2.2. The Basic Principles of Greedy Algorithm

In the past few years, iterative greed has been used to solve quite a few problems [9]. A greedy algorithm is a commonly used method for finding the optimal solution to problems. A Method of Finding Local Optimal Solution and Obtaining Global Solution [10]. This method pattern generally divides the solving process into several steps, but each step applies the principle of greed, selecting the best/optimal choice in the current state (the locally most advantageous choice), and hoping that the final stacked result is also the best/optimal solution.

#### 2.2.1 Establish replenishment and pricing strategies for various categories of vegetables in the coming week.

Firstly, we processed the survey data from Google and SCI-Hub to obtain the total sales of each vegetable category and its products. Considering that supermarkets need to make replenishment plans based on categories, but supermarkets prioritize profitability, we need to comprehensively consider the relationship between total sales and cost-plus pricing, and make reasonable replenishment plans for each vegetable category to maximize supermarket revenue. And provide pricing strategies for the total daily replenishment amount of each vegetable category in the next week (July 1-13, 2023). The greedy algorithm flow chart is shown in Figure 1.

The greedy algorithm is solved by constructing a two-layer greedy that first satisfies the outer greedy: the "most economical" ordering solution, and second solves the inner greedy: the least transportation loss. Then we use the distance to optimal solution (TOPSIS) method to solve the conditions of the "most economical" ordering solution and least transportation loss. Since it is assumed that the maximum number of items that can be replenished in the sales space is no more than 50, the top 50 vegetable items are selected by satisfying these two sets of constraints.

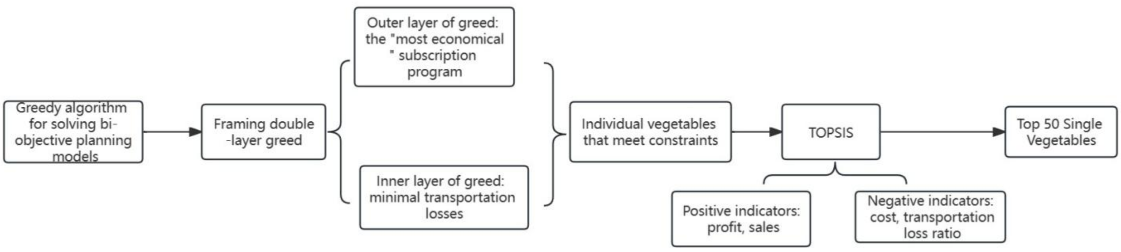


**Figure 1** Flowchart of the greedy algorithm

## 3. Results

### 3.1. Data preprocessing

Considering that superstores have to make replenishment plans on a category-by-category basis, but supermarkets are profit-oriented, we have to consider the relationship between total sales and cost-plus pricing, and make reasonable replenishment plans for each vegetable category, so as to maximize the superstore's revenue. The pricing strategy for the total daily replenishment for the coming week (July 1-13, 2023) for each vegetable category is also given.

When processing the data, we found that some of the data had codes for local spinach and lotus root, but some of them did not have records of purchase and sale of local spinach and lotus root by supermarkets, so we used the exclusion process for such cases to avoid the reduction of the effective mean caused by the subsequent calculation of the average value under the apportionment. At the end of data processing, we found the existence of the same single product vegetables from different suppliers, resulting in the same vegetables have repeated problems, we repeated the same vegetables and then go to the product mean value as the final results to calculate.

The required objective function is then calculated, and in calculating the sales pricing, we select the annexed plurality of the pricing of that single vegetable, which is the undiscounted pricing, to be used as the sales pricing of the single item. The cost-plus rate after discounting was then analyzed using linear regression (least squares), i.e., analyzing the coefficients of the wholesale price (￥/kg) and the unit sales price (￥/kg).

Table 1 uses the least squares method to analyze the fitted regression equations and fitted regression values of cost markup rates for six different categories of discounts. Among them, $R^2$ Indicates the fitting degree of curve regression, and the closer it is to 1, the better the effect. Then calculate the average wholesale price between different categories and bring it into the fitted regression line to obtain the sales price. Subtract the wholesale price to obtain the profit situation of each category at a discount, which is convenient for subsequent profit calculation.

**Table 1:** Cost markup rate fitting regression table for six categories

| Category | formula | R2 | Fit degree | Average wholesale price | Discounted profit |
|---|---|---|---|---|---|
| florescent vegetables | $y = 0.703 + 1.085x_1$ | 0.521 | Strong | 6.077 | 1.213 |
| Florifolias | $y = 2.682 + 0.435x_2$ | 0.382 | Weak | 3.568 | 0.667 |
| Chili peppers | $y = 1.067 + 1.083x_3$ | 0.702 | Strong | 5.084 | 1.489 |
| Solanaceae | $y = 2.855 + 0.787x_4$ | 0.35 | Weak | 4.552 | 1.886 |
| edible fungi | $y = 0.051 + 1.176x_5$ | 0.802 | strong | 5.600 | 1.037 |
| Aquatic rhizomes | $y = 2.895 + 0.826x_6$ | 0.504 | strong | 9.712 | 1.205 |

### 3.2. The predicted result of replenishment strategy

During the solving process, we check whether at least one product has been selected for each category and record the selected categories and quantities. Then, we select the remaining unselected products in ascending order of transportation loss rate until at least one product is selected for each category, or the cumulative number of selected products reaches 50.

**Table 2:** Top 50 vegetable items

| Item Name | Classification Name | Profit Operating | Loss Rate |
|---|---|---|---|
| Seven colored peppers | Chili peppers | 220.4398 | 0.0924 |
| Luffa tip | Florifolias | 106.7109 | 0.1283 |
| Ice grass | Florifolias | 64.09539 | 0.1283 |
| Agaricus Bosporus | edible fungi | 94.89179 | 0.0945 |
| Sichuan Red Toon | Florifolias | 249.1244 | 0.1283 |
| Echinacea officinalis | Florifolias | 73.2513 | 0.1283 |
| Jigu mushroom | edible fungi | 84.64055 | 0.0945 |
| Chinese cabbage | Florifolias | 77.07525 | 0.1283 |
| Xiaomi pepper | Chili peppers | 114.508 | 0.0924 |
| Sophora japonica | Florifolias | 61.21581 | 0.1283 |
| Fruit chili | Chili peppers | 188.4801 | 0.0924 |
| Hongshan Cabbage Stalk Rare Handbag | Florifolias | 335.0118 | 0.1283 |
| …… | …… | …… | …… |

Use seasonal ARIMA to calculate the daily replenishment total and pricing strategies for each category of vegetables in the next week.

Figure 2-3 below shows the results of the sales and pricing strategy projections for each of the two categories of vegetables for the coming week.


**Figure 2** Ginger sales and pricing forecast results


**Figure 3** Garlic sales and pricing forecast results

As shown in Figure 2 and 3, for the storable vegetable varieties such as ginger and garlic, but their prices have a long-term trend of rotation. Through consulting the data, it is found that in the production, storage and circulation links, the asymmetry of market information and the imperfect circulation system increase the possibility of malicious hoarding and price manipulation by speculators. Therefore, the government should improve the market information transmission mechanism, timely release authoritative market information, improve the transparency of the vegetable market, ensure the balanced supply capacity of the market, and prevent price spikes and crashes.

### 3.3. The results of the optimization model

This paper establishes an optimization model about profit maximization, and uses greedy algorithm to solve it, so as to obtain supermarket replenishment model under profit maximization. According to our analysis, we take profit maximization as the goal, Then, we established the following equations with the constraints of maximizing profit and minimizing shipping loss rate, the decision variables of replenishment and shipping rate, and the objective functions of cost, cost plus rate, pricing, sales revenue, profit, and vegetable surplus,The specific optimization model is as follows:

$$\max \sum_{i=1}^{6} R_i - C$$

$$s.t.\begin{cases} C = \sum_{i=1}^{6} p_{i,k} d_{i,k} \\ s_{i,k} = C_{i,k}(1+r_i) \\ R_i = \sum_{i=1}^{6} (1-l_k) d_{i,k} \cdot s_{i,k} + \sum_{i=1}^{6} l_k \cdot d_{i,k} \cdot s_{i,k}' \end{cases} \tag{3}$$

In summary, there is a certain correlation between the total sales volume and cost-plus pricing of various vegetable categories, with garlics having the weakest correlation and Leek flower having the strongest. Garlics, edible mushrooms, and aquatic rhizome vegetables will have a positive impact on the total sales volume after discounts, while the higher the markup cost pricing before discounts, the lower the sales volume. On the contrary, vegetables such as Leek flower, leaves, and chili peppers harm total sales after discounts.

Figures 4-5 show the results of sales and pricing forecasts for each category of vegetables, respectively.

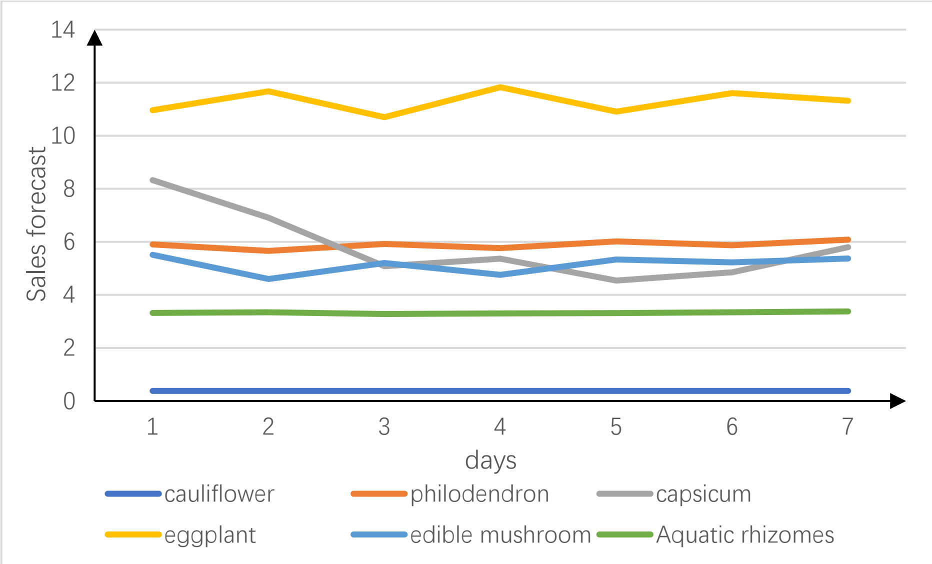


**Figure 4** Results of the sales forecast for the next 7 days for each category of vegetables

From the data in Figure 4, the sales volume of different categories of vegetables may vary over the next 7 days, but the overall trend is relatively stable. Sales in the cauliflower category remained stable over the seven-day period, at 0.381 per day. sales in the foliage category increased incrementally from day to day, starting at 5.9 and increasing by about 0.1 per day. sales in the pepper category declined and then increased, with a low of 5.084 on the third day, and then began to pick up on the fourth day. Sales of eggplant increased incrementally for the first four days, dropped slightly on the fifth day, and then increased incrementally through the seventh day. Sales of edible mushrooms fluctuated, but the overall trend was relatively smooth. Sales of aquatic roots and tubers were also more stable, at around 3.3 per day.

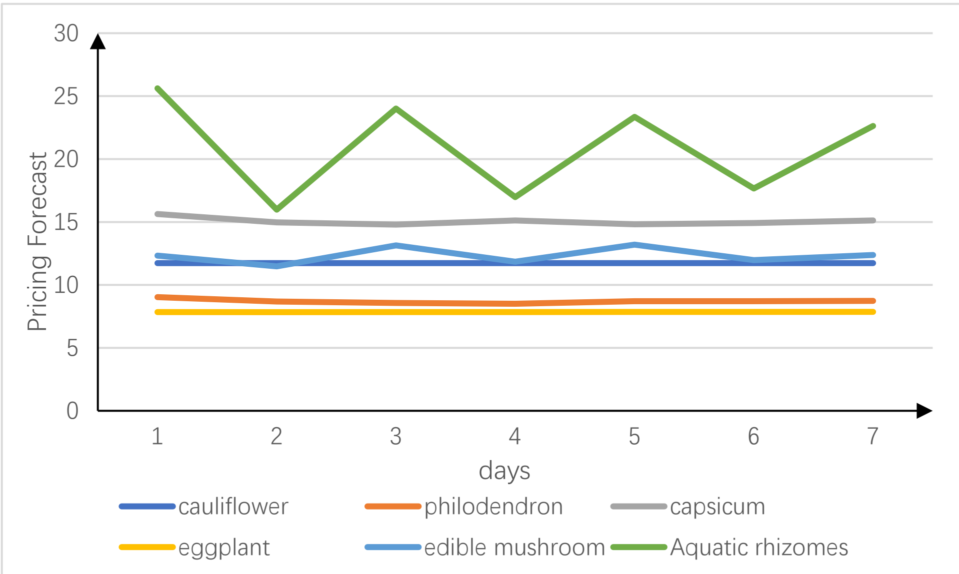


**Figure 5** Pricing Forecast Results by Vegetable Category

From the data in Figure 5, it can be seen that the pricing forecasts for various vegetables vary over the next seven days. The cauliflower category has the same pricing for each day at 11.7331. the leafy floral category has a slight change in pricing from 9.031 incrementally to 9.073. the pepper category has an incremental increase in pricing each day from 14.793 to 15.127, then a decremental decrease to 14.815, and finally an incremental increase to 15.116. the eggplant category has an incremental increase in pricing each day from 7.837 to 7.861. the edible mushrooms category Pricing fluctuated daily, increasing to 13.2, then decreasing to 11.975, and finally increasing to 12.378.

Daily pricing for aquatic roots and tubers increased to 25.619, then decreased to 15.981, then increased to 24.012, then decreased to 16.978, and finally increased to 23.356.

For these categories of vegetables, future sales volumes and pricing will be affected by a number of factors, including market demand, seasonal variations, price fluctuations and climate change. Specific sales volume forecasts need to take into account the combined effects of these factors. In addition, the data in Figures 4-5 may be derived from certain trend analysis or time series forecasts based on historical data, and the specific sales volume results need to be judged and adjusted according to the actual situation and the latest market conditions.

## 4. Conclusions

Based on the above analysis, there are differences in sales and pricing predictions for different vegetable categories. Supermarket stores need to develop corresponding replenishment and pricing strategies based on market demand, competition, and cost considerations for different categories, to increase revenue and meet market demand.

This article delves into and analyzes the application of time series models and greedy algorithms in mathematical modeling. By modeling and analyzing time series models, researchers can extract useful information from the data of supermarkets in the past three years to analyze the relationships between various categories and individual products. On the other hand, the greedy algorithm, as a commonly used optimization algorithm, gradually obtains the global optimal solution by selecting a local optimal solution at each stage. The experimental results show that the dual objective greedy algorithm optimization model has good efficiency and applicability and has certain practical application value.

## References

[1] Bhilwadikar T, Pounraj S, Manivannan S, et al. Decontamination of microorganisms and pesticides from fresh fruits and vegetables: A comprehensive review from common household processes to modern techniques [J]. Comprehensive reviews in food science and food safety, 2019, 18(4): 1003-1038.

[2] Wang X. Joint Decision Making of Replenishment, Pricing, and Fresh Keeping Input in Fruit and Vegetable Cold Chain: Based on Markov Process [J]. Mobile Information Systems, 2022, 2022.

[3] Tashakkor N, Mirmohammadi S H, Iranpoor M. Joint optimization of dynamic pricing and replenishment cycle considering variable non-instantaneous deterioration and stock-dependent demand [J]. Computers & Industrial Engineering, 2018, 123: 232-241.

[4] Mahmoodi A. Joint pricing and inventory control of duopoly retailers with deteriorating items and linear demand [J]. Computers & Industrial Engineering, 2019, 132: 36-46.

[5] Fan T, Xu C, Tao F. Dynamic pricing and replenishment policy for fresh produce [J]. Computers & Industrial Engineering, 2020, 139: 106127.

[6] Duan Y, Liu J. Optimal dynamic pricing for perishable foods with quality and quantity deteriorating simultaneously under reference price effects [J]. International Journal of Systems Science: Operations & Logistics, 2019, 6(4): 346-355.

[7] Adhikari R, Agrawal R K. An introductory study on time series modeling and forecasting [J]. ar**v preprint ar**v:1302.6613, 2013.

[8] Siami-Namini S, Tavakoli N, Namin A S. A comparison of ARIMA and LSTM in forecasting time series [C]//2018 17th IEEE international conference on machine learning and applications (ICMLA). IEEE, 2018: 1394-1401.

[9] Rodriguez F J, Lozano M, Blum C, et al. An iterated greedy algorithm for the large-scale unrelated parallel machines scheduling problem[J]. Computers & operations research, 2013, 40(7): 1829-1841.

[10] Deng S, **ong H, Yang Q, et al. On Enhancing Transmission Performance for IoV Based on Improved Greedy Algorithm [C]//International Conference on Edge Computing and IoT. Cham: Springer Nature Switzerland, 2022: 41-53.