A Decision Model for Pricing and Replenishment of Vegetable Products Based on Lagrangian Function Optimization

Huaping Mei *, Jiarong Li, Chao Pan

School of Building Services Science and Engineering, Xi'an University of Architecture and Technology, Xi'an, China, 710055

* Corresponding Author Email: MHP15971651897@outlook.com

Abstract. In the era of big data, superstores usually replenish and price their products based on the historical sales and demand of each item. Based on data mining, this paper analyzes the correlation between different categories using a gray correlation analysis model, then fits the relationship between total sales and cost-plus pricing for each vegetable category, and uses a time series forecasting model to predict the wholesale price and sales volume of six types of vegetables in the week of July 1st to 7th. By establishing a Lagrangian function-based optimization model with the objective function of maximizing the profit of the superstore, the total daily replenishment of each category and the cost-plus pricing as the decision variables, and traversing the solution with a step size of 0.5 sales, the total daily replenishment and the pricing strategy for the next seven days can be formulated, and the total return of the superstore in the coming week is found to be 2264.41 yuan. The optimization model based on profit maximization of superstores used in this paper provides reliable theoretical support and practical guidance for merchants to predict market conditions and plan future replenishment and pricing strategies.

Keywords: Gray Correlation Analysis Model, Time Series Forecasting, Lagrangian Function Based Optimization Model, Iterative Method.

1. Introduction

The general freshness period of vegetable commodities is relatively short, when the sales time increases, the quality deteriorates, resulting in most varieties can not be re-sold the next day[1] . From 3:00am to 4:00am is generally the time of the day when vegetables are purchased and traded. Due to the inaccuracy of specific items and prices, merchants are required to make replenishment decisions for each vegetable category based on the usual sales and demand for each item. Vegetables are priced using the "cost-plus pricing"[2] method, and supermarkets often sell at a discount for damaged and deteriorating products. On the demand side, there is a correlation between the sales volume and the time of the year, and on the supply side, there are more varieties of vegetables available from April to October each year. Therefore, it is important to explore and analyze the stable market demand relationship for the automatic pricing and replenishment decision of commodities.

With the development of operations research, in modern scientific decision-making, often with the help of natural science methods, the use of mathematical tools, the establishment of the relationship between the decision-making variables of the formula and model, to reflect the essence of the decision-making problem, to simplify the complex decision-making problem. In this paper, the optimization model based on the Lagrangian function can help superstores to formulate replenishment and pricing strategies without exactly knowing the specific single product and purchase price, and help superstores to maximize the total profit[3] .

2. Modeling

2.1. Gray correlation analysis modeling

Gray correlation analysis is a multi-factor comprehensive evaluation method for studying the degree of association between multiple factors. Its basic principle is to quantify the correlation between factors into a gray correlation value by constructing a gray correlation model, so as to realize the comparison and ranking of the degree of correlation between factors.

The basic steps of gray correlation analysis are as follows:

Step1: Determine the object of evaluation

Determining the evaluation object means determining the parent and subsequence. One of the six categories is selected as the parent sequence, and the other five are sub-sequences. Similarly, one of the 35 individual items is selected as the parent sequence, and the other items are sub-sequences[4] .

Step2: Indicator Positive

The indicator variable Vegetables is positively normalized across categories and individual sales, and therefore does not need to be normalized.

Step3: Construct the matrix

Convert sequence data to a gray correlation matrix.

Step4: Standardization

Normalization of the normalized matrix was used to eliminate the scale. For each indicator in the parent series and the subseries, the mean value of each indicator was obtained and each element of the indicator was divided by its mean value.

Step5: Calculate the gray correlation coefficient

The association coefficients of the corresponding elements of each comparison sequence are computed separately.

Step6: Calculate the gray correlation

Mean each column of the correlation coefficient matrix.

The flowchart of the gray correlation analysis algorithm can be represented as figure 1:

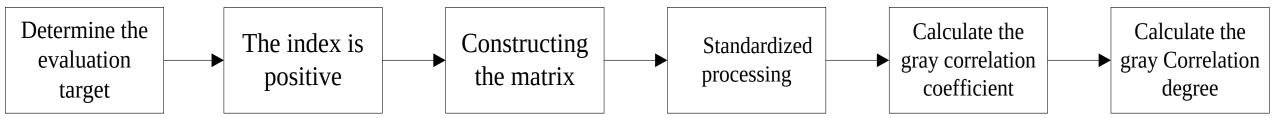


Figure 1 Flowchart of gray correlation analysis algorithm

2.2. Time series forecasting modeling

ARIMA(p,d,q) is a differential integrated moving average autoregressive model that utilizes past observations of the series to extrapolate future values. Where p is the order of the autoregressive part, d is the number of difference operations and q is the order of the moving average part[5] .

Step1: Extract time series data

The wholesale prices and sales volumes of six categories of vegetables were processed as one-dimensional vectors.

Step2: Smoothness test of time series

A non-smooth time series is transformed into a smooth time series by means of a difference transform or a logarithmic difference transform.

Step3: Model Ordering

After determining the difference order d, the autoregressive order p and sliding average order q are confirmed based on the ACF plot, PAC plot, AIC criterion, and BIC criterion, and the best model is selected.

Step4: Parameter Estimation and Diagnostic Tests

Based on the model parameters such as the number of samples, degrees of freedom, Q-statistics and the information criterion model's goodness of fit R^2 , the model effect was tested.

Step5: Model Prediction

The established ARIMA model was used to forecast the wholesale price and sales volume of each vegetable category for the next seven days.

The time series forecasting flowchart can be represented as figure 2:

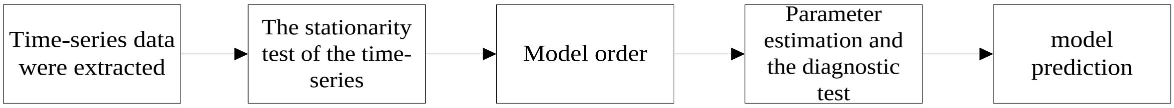


Figure 2 Flowchart of time series forecasting

3. Empirical studies

3.1. Data pre-processing

The existing data information will be used to complete the product information and integrate the data through pivot table. Pivot the sales of categories and individual products by year, quarter, month, day, time period, etc.[6] . Combined with the title sweep sales start and end time and the actual situation of life, the time period is quantified, so that before 12:00 is 0, 12:00-18:30 is 1, and after 18:30 is 2. Then the box-and-line diagram is used to monitor the outliers of total sales of flowers and foliage categories and the cost-plus rate of each category and exclude them in order to ensure the accuracy of the model. The results of the obtained data preprocessing are shown below figure 3:

Sum: sale	Column labels						
Row labels	Cauliflower	Mosaic and leafy	Chili peppers	Nightshades	edible fungi	Aquatic rhizomes	total
In 2020	9399.216	39167.661	12149.233	4776.843	14604.582	6177.806	86275.341
Season 3	4593.107	19438.689	6990.856	3166.152	4877.197	1990.977	41056.978
Season 4	4806.109	19728.972	5158.377	1610.691	9727.385	4186.829	45218.363
October	1635.955	6359.745	1901.931	770.409	2991.254	1339.933	14999.227
November	1938.822	6487.149	1832.441	431.825	3270.466	956.862	14917.565
December	1231.332	6882.078	1424.005	408.457	3465.665	1890.034	15301.571
In 2021	12646.78	58496.536	23158.267	7767.475	21849.729	13071.615	136990.402
In 2022	14119.578	66898.77	34569.072	6239.996	23263.506	15704.803	160795.725

Figure 3 Category pivot table

3.2. Distribution pattern of each category

Using a pivot table, the average monthly sales for four years can be found. According to the results can be derived from the largest sales of flowers and foliage. May-August average monthly sales showed a growing trend, combined with the actual life can be seen, the time period in the ripening and harvesting of fruits and vegetables, supermarkets wholesale volume, corresponding to the sales of vegetables, while corresponding to the question of the supply of vegetables in the question in the April-October varieties of the more abundant.

Let each day before 12:00 be recorded as morning sales, 12:00-18:30 as afternoon sales, and after 18:30 as evening sales. Visualizing the data of categories stacked in bar charts with percentage of sales by different time periods in a day, it was concluded that the afternoon sales of cauliflower, cauliflower, pepper, eggplant, and edible mushrooms accounted for the largest portion of the day's sales, which may be due to the fact that during the afternoon hours of the day, the quality of the dishes gradually declined, and therefore they were sold at a discount, thus increasing the sales volume,as shown in Figure 4.

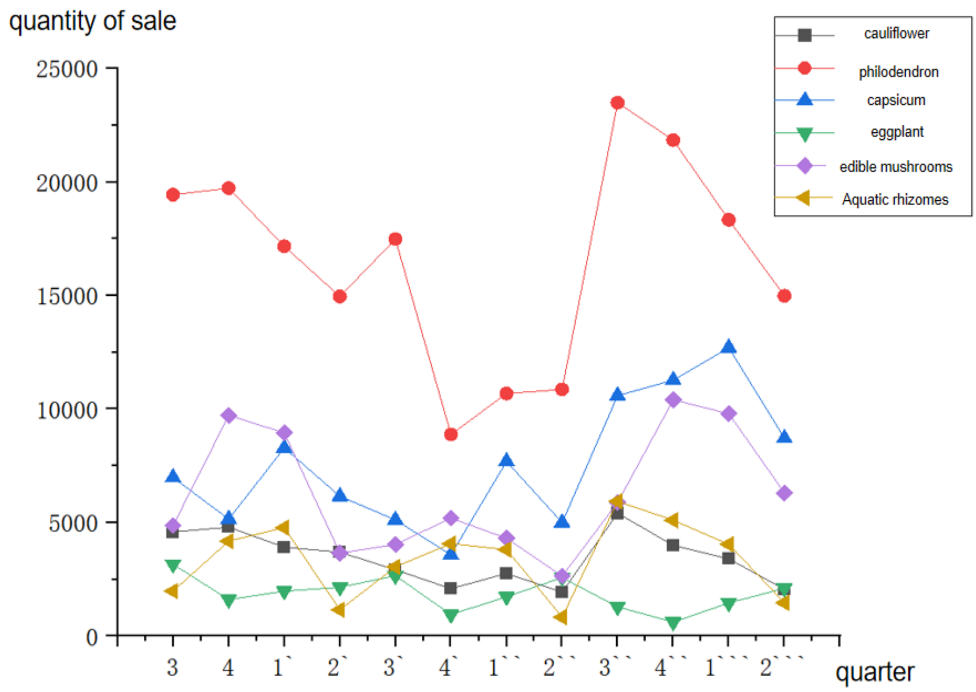


Figure 4 Line graph of sales volume by category in different quarters

3.3. Solving the gray correlation analysis model

After selecting any one of the 6 categories as the parent sequence, the other 5 items are the sub-sequences. After the gray correlation analysis of the 6 categories in turn, the gray correlations in the following table 1 can be derived:

Table 1 Category gray correlation table

	cauliflower	philodendron	capsicum	edible mushroom	Aquatic rhizomes	eggplant
cauliflower	1	0.764	0.647	0.599	0.59	0.579
philodendron	0.797	1	0.703	0.594	0.59	0.62
capsicum	0.684	0.704	1	0.631	0.591	0.588
Edible mushroom	0.673	0.639	0.67	1	0.707	0.532
Aquatic rhizomes	0.683	0.657	0.651	0.728	1	0.522
eggplant	0.676	0.682	0.654	0.654	0.523	1

From the above table 1, it can be seen that the largest correlation in the category of cauliflower is the category of flowers and leaves, with a gray correlation of 0.764; at the same time, it is found that the largest correlation in the category of flowers and leaves is the category of cauliflower, with a magnitude of 0.797. The two are the largest party of the gray correlation of the other side of the other side of the other side of the other side of the other side of the other side of the biggest party of the other side of the gray correlation can be drawn that the two have a strong correlation, and with the correlation law of this correlation, there is also the category of edible mushrooms and aquatic roots. The above table is visualized as shown below figure 5.

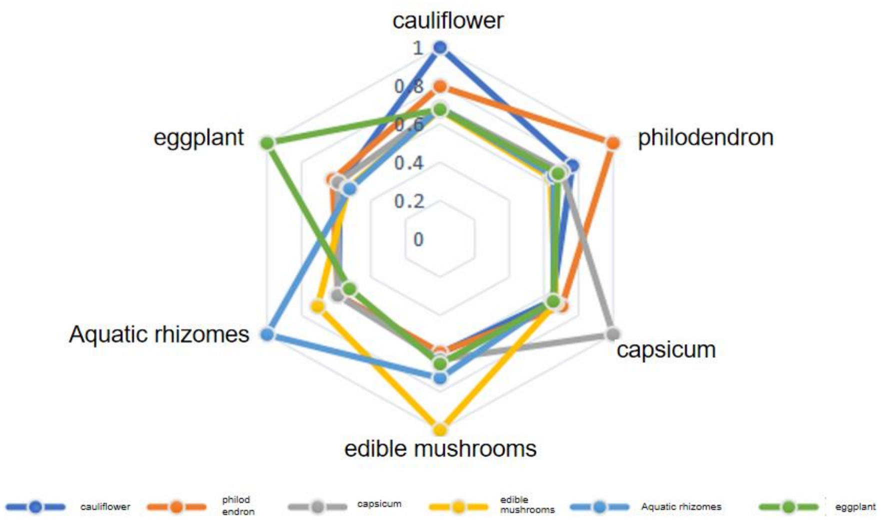


Figure 5 Gray correlation radar chart

The results of the Kendall coefficient of concordance test showed that the overall data showed significance at the level of P-value of significance less than 0.001, which presents significance and rejects the original hypothesis, therefore the data shows consistency. Also the Kendall's coefficient of concordance W value of the model is 0.898, so the degree of correlation is high.

3.4. Develop total daily replenishment and pricing strategies

Outliers were first monitored and eliminated using box-and-line plots of total sales in the foliage category versus the cost-plus ratio for each category as figure 6.

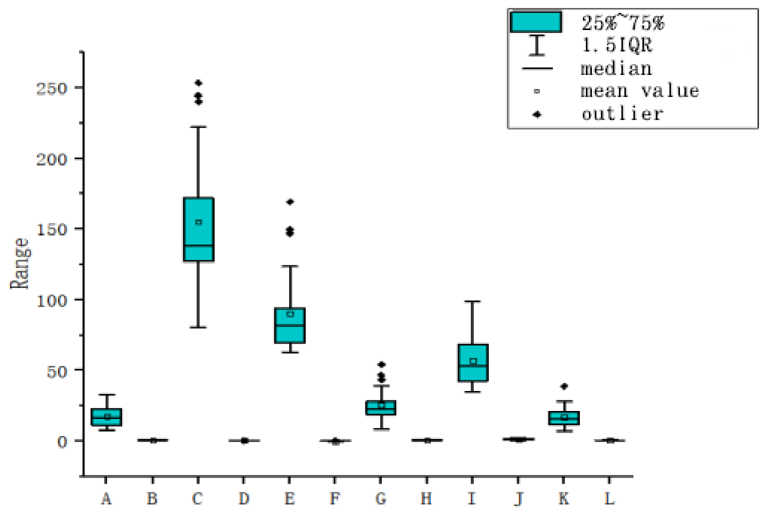
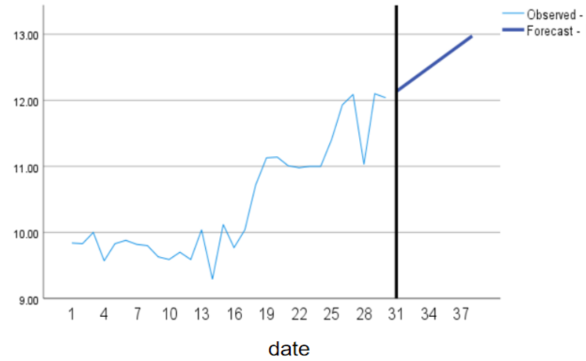


Figure 6 Box-and-line diagram

The fitted values are derived from the historical true values and projected for the next seven days. A time series plot of wholesale cost and sales of aquatic roots and tubers for the next seven days can be derived as figure 7 and figure 8.



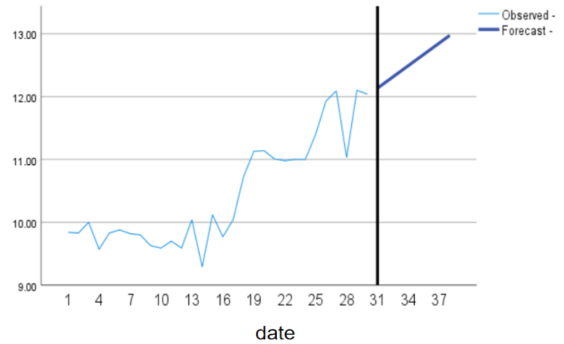


Figure 7 Projected wholesale cost of aquatic roots and tubers **Figure 8** Forecasted sales of aquatic roots and tubers

The forecasted wholesale cost results for the week of July 1-7 for six categories of vegetables are as table 2.

Table 2 Forecast of wholesale costs for the category for the coming week

dates	cauliflower	philodendron	capsicum	eggplant	edible mushroom	Aquatic rhizomes
July 1	8.27	3.65	6.25	4.48	4.68	12.05
July 2	8.49	3.73	6.26	4.60	4.81	12.13
July 3	8.59	3.78	6.26	4.70	4.72	12.21
July 4	8.63	3.81	6.26	4.78	4.61	12.28
July 5	8.65	3.83	6.27	4.84	4.63	12.36
July 6	8.66	3.84	6.27	4.88	4.65	12.43
July 7	8.67	3.84	6.27	4.90	4.60	12.51

The results of predicting the sales of six categories of vegetables for the week of July 1-7 are as table 3.

Table 3 Forecast of category sales for the coming week

dates	cauliflower	philodendron	capsicum	eggplant	edible mushroom	Aquatic rhizomes
July 1	12.05	4.94	7.28	8.47	11.51	16.03
July 2	12.48	4.92	7.06	8.44	11.6	16.08
July 3	12.81	4.89	7.05	8.41	11.7	16.13
July 4	13.05	4.87	7.23	8.38	11.79	16.18
July 5	13.23	4.84	7.19	8.35	11.89	16.23
July 6	13.37	4.82	7.07	8.31	11.97	16.28
July 7	13.47	4.80	7.12	8.28	12.07	16.36

The two tables above utilize a time series forecasting model to predict the sales volume and wholesale cost of each category of vegetables for the coming week. Since there is a correlation between the sales volume and cost-plus pricing of each category of vegetables, selecting the appropriate cost-plus pricing for each category for each day of the coming week has an important impact on the profitability of the superstore.

3.5. Optimization model based on Lagrangian function

Lagrange multiplier method, as an optimization algorithm, Lagrange multiplier method is mainly used to solve constrained optimization problems, its basic idea is to transform a constrained optimization problem with n variables and k constraints into an unconstrained optimization problem with (n+k) variables by introducing Lagrange multipliers. The mathematical meaning behind the

Lagrange multipliers is that they are the coefficients of each vector in a linear combination of the gradients of the constraint equations.

In the second half of Problem 2, the total daily replenishment and pricing strategy for the vegetable category for the coming week need to be given. For this purpose, the total sales volume of each individual item under each category represents the sales volume of the category, and the average of the cost and sales price of each individual item under each category represents the wholesale and sales price of this category. There is a certain relationship between the sales volume of each vegetable category and the cost plus price, in order to get the market situation in the coming week, combined with the historical data of June 2023, a total of thirty days, the use of time-series models to predict the wholesale price and sales volume of six types of vegetables in the week of July 1st to 7th. Finally, an optimization model[7]is established with the objective function of maximizing the profit of the superstore, and the total daily replenishment of each category and cost-plus pricing as the decision variables.

Let x_{ij} denote the markup cost pricing on day i for the j category. w_{ij} Denote the interest rate of day i for category j . W_{ij} Denote the profit on day for category i j . Y_{ij} Indicates the total sales volume of the category on day j i c_{ij} . Indicates the cost of the category on day j i δ_{ij} .denotes the shipping loss rate of the j category on day i . $S_{j\min}$ and $S_{j\max}$ represent the minimum and maximum historical sales values for category j , respectively.

Establish an objective function in terms of maximizing the profit of the superstore:

$$max\sum W = \sum_{i=1}^{7}\sum_{j=1}^{6} c_{ij} w_{ij} Y_{ij} (1-\delta_{ij}) \tag{1}$$

Constraints:

(1) There is a relationship between sales volume and markup cost pricing for each category:

$$Y_{ij} = f(x_{ij}) \tag{2}$$

(2) The equation for the relationship between cost-plus pricing and costs as well as interest rates:

$$X_{ij} = c_{ij}(1+w_{ij}) \tag{3}$$

(3) According to statistical principles, the sales of each category should lie between the maximum and minimum values of historical sales:

$$S_{j\min} \le Y_{ij} \le S_{j\max} \tag{4}$$

(4) $X_{ij} > 0, Y_{ij} > 0, w_{ij} > 0.\ x_{ij}, Y_{ij}, w_{ij} \in R\ (i = 1, 2 \ldots 7,\ j = 1, 2 \ldots 6)$

Combining the above equations, the objective function can be summarized to maximize the revenue of the superstore and develop a planning model for the total daily replenishment and pricing strategy:

$$max\sum W = \sum_{i=1}^{7}\sum_{j=1}^{6} c_{ij} w_{ij} Y_{ij} (1-\delta_{ij}) \tag{5}$$

$$\text{S.t.} \begin{cases} Y_{ij} = f(x_{ij}) \\ X_{ij} = c_{ij}(1+w_{ij}) \\ X_{ij} > 0, Y_{ij} > 0, w_{ij} > 0 \\ S_{j\min} \le Y_{ij} \le S_{j\max} \\ x_{ij}, Y_{ij}, w_{ij} \in R \end{cases} \tag{6}$$

3.6. Solution of the traversal algorithm

Traversal means that each node in a tree (or graph) is visited once along a certain search route. The operations performed to access a node depend on the specific application problem, and the specific access operations may be checking the value of a node, updating the value of a node, etc.[8] . The order in which the nodes are accessed is different for different traversal methods.

To solve the above optimization model, the replenishment plan that maximizes the total profit is found by iterating through all possible combinations of replenishment quantities, given the range of replenishment quantities for each type of commodity. Specifically, the MATLAB code employs a nested loop structure that traverses the range of replenishment quantities for each type of commodity in turn. In each loop, the replenishment quantities are filtered based on certain conditions, and then the corresponding margins are calculated and the total profit for that combination is computed. If the total profit for that combination is greater than the current maximum total profit, the maximum total profit is updated and the corresponding replenishment quantity and profit margin are recorded. The following traversal algorithm solution flowchart is derived as figure 9.

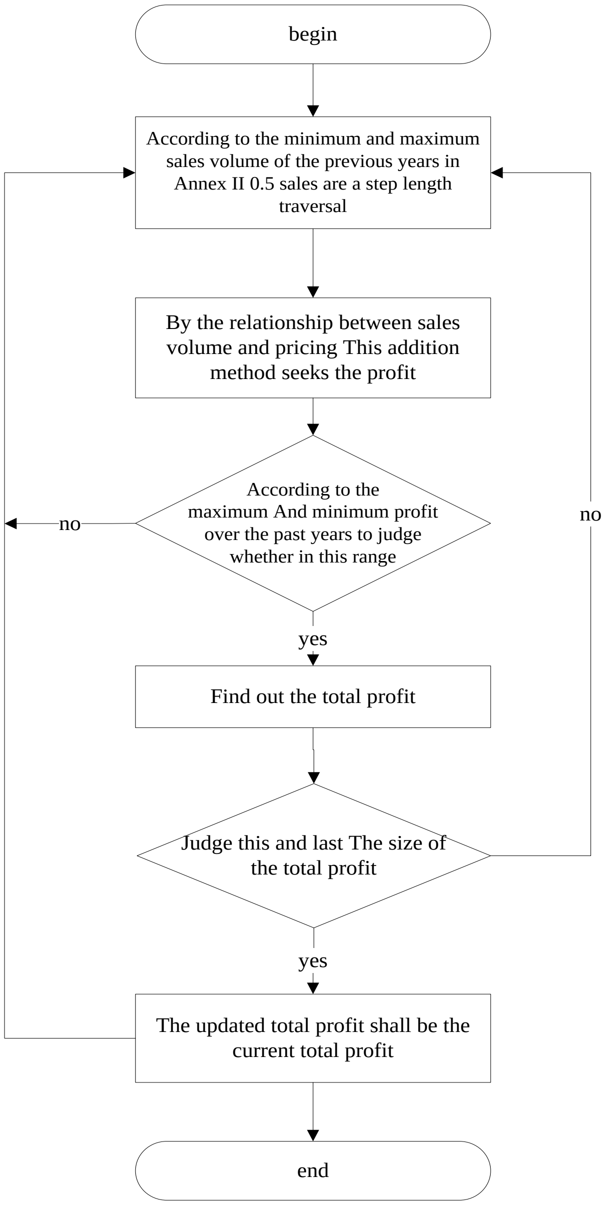


Figure 9 Flowchart of the traversal algorithm solution

The above algorithmic flowchart, after solving, leads to the total daily replenishment scenarios and pricing strategies for July 1-7, 2023, for each vegetable category as shown in the table 4 and table 5.

Table 4 Total daily replenishment program

dates	cauliflower	philodendron	capsicum	eggplant	edible mushroom	Aquatic rhizomes
July 1	22.300	185.448	85.266	20.500	58.912	16.526
July 2	21.250	183.798	85.216	20.550	58.512	16.426
July 3	20.750	182.748	85.216	20.600	58.762	16.326
July 4	20.550	182.148	85.216	20.600	59.112	16.226
July 5	20.450	181.698	85.216	20.650	59.062	16.126
July 6	20.400	181.498	85.216	20.650	59.012	16.076
July 7	20.350	181.498	85.216	20.650	59.162	15.976

Table 5 Pricing Strategy

dates	cauliflower	philodendron	capsicum	eggplant	edible mushroom	Aquatic rhizomes
July 1	15.59	6.00	9.43	8.79	13.86	18.61
July 2	15.99	6.13	9.46	9.02	13.92	18.65
July 3	16.18	6.21	9.46	9.25	13.88	18.69
July 4	16.25	6.26	9.46	9.25	13.83	18.73
July 5	16.29	6.30	9.46	9.49	13.83	18.77
July 6	16.31	6.31	9.46	9.49	13.84	18.79
July 7	16.33	6.31	9.46	9.49	13.82	18.83

Under the above decision scheme, the sum of the product of the total daily replenishment and the profit for these six categories is the return for the next seven days as table 6.

Table 6 Proceeds

dates	1	2	3	4	5	6	7
earnings volume	326.63	319.98	322.19	324.85	324.02	322.59	324.14

4. Discussion of results

After analyzing the correlation between different categories using the grey correlation analysis model, Kendall's coefficient in statistics was introduced to test the grey correlation analysis model, which made the accuracy of the conclusions much higher.

Before formulating the total daily replenishment and pricing strategy for the coming week, a time series model is used to predict the wholesale price and sales volume of different categories for the coming week based on the historical data of each category of vegetables, which makes the establishment of the optimization model later have more data support. The statistical model used in this paper is based on the historical data of each category and individual vegetables, which enables merchants to make reasonable decisions in the face of future uncertainties.

Given the limited data given and known conditions, in order to better formulate the replenishment and pricing decisions for vegetable commodities, in addition to the historical wholesale price, sales price, sales volume, and shipping loss rate of each individual vegetable, the superstore should also consider the degree of influence of national policy orientation, natural disasters, epidemics, and other uncertainties[9] on the market sales quotations of vegetable commodities, and fully grasp the future market conditions as well as the external marketing environment.

From the supply side and demand side, it is also necessary to collect relevant data such as consumer satisfaction, the operation of peers around the superstore, the initial loss situation when the supply chain transports vegetables to the superstore, the time when the quality of different dishes changes, the differences in the preferences of local consumers for different single products and categories of dishes, and the statistical survey on the age structure of the consumer population[10] , etc., so that the replenishment of vegetable commodities and pricing decisions can be better formulated. The data can be used to make better decisions on replenishment and pricing of vegetable products.

5. Conclusions

After data preprocessing, the correlation between different categories and individual products was analyzed using the gray correlation analysis model: the consistency of the gray correlation between cauliflower and foliage categories was obtained, and it can be concluded that the two have a very strong correlation. There are also edible mushrooms and aquatic roots and tubers with this correlation pattern. Then, the gray correlation analysis model was tested by Kendall coefficient consistency test, and it was concluded that the Kendall coordination coefficient W value of the model was 0.898, and the model was effective.

Using the time series prediction model to derive the wholesale price and sales volume of six categories of vegetables in the week from July 1 to 7, the optimization model based on the Lagrangian function is established with the maximal profit of the superstore as the objective function, and the total daily replenishment quantity of each category and the cost-plus pricing as the decision variables. According to the minimum and maximum values of the sales volume in the data for the calendar year, the total daily replenishment and pricing strategy for the next seven days are formulated by traversing the solution in steps of 0.5 sales volume, and the total return of the superstore in the coming week is found to be $2,264.41, with the maximum returns of $326.63, $319.98, $322.20, $324.85, $324.02, and $322.59, respectively, 324.14 dollars.

The optimization model used in this paper based on profit maximization under superstores is of deep practical significance and can be extended to a variety of merchant ordering and pricing problems. At the same time, the statistical model used in this paper is based on historical data of various categories and single vegetables, which enables merchants to make reasonable decisions in the face of future uncertainties.

The time series forecasting model used in this paper can extract the long-term, seasonal, cyclic and irregular trends of the data, which can be used to make economic development forecasts or transaction forecasts in social production, such as forecasting the sales of cold beverages, forecasting the sales of clothing, and forecasting the weather temperature.

References

[1] Cui Ligang, Li Yali, Liu Jinxing et al. Joint decision making for multi-product replenishment and pricing considering investment in preservation technology [J]. Industrial Engineering and Management, 2023, 28(03): 17-26.

[2] Lin Xiaodong. Analysis of the disadvantages of cost-plus pricing method and its reform [J]. Introduction and Consultation, 2005(11): 13-14.

[3] YANG Tianshan,YUAN Gonglin. Research on dynamic pricing and replenishment strategy of perishables considering green distribution technology investment [J/OL]. China Management Science: 1-13.

[4] Qiyuan Jiang, Jinxing Xie, Jun Ye. Mathematical modeling [M]. Beijing: Higher Education Press. 2018.

[5] Shoukui Si, Zhaoliang Sun, Xijing Sun. Mathematical modeling algorithms and applications [M]. National Defense Industry Press, 2015.

[6] Shen Limin, Li Jian. A data processing method based on Power Query and pivot table [J]. Computer Knowledge and Technology, 2021, 17(36): 25-27.

[7] TIAN Junfeng, SUN Xixiu, YANG Mei. A generalized model for joint decision making of replenishment and pricing considering demand and price-quality correlation [J]. Logistics Engineering and Management, 2014, 36(05): 188-192.

[8] HE Jinming, LIU Jianglong, NIU Lei. Traversal algorithm based minimum spacing solution for complex overhead transmission lines [J]. Electric Power Survey and Design, 2022(07): 45-47+75.

[9] Ma Jiawei. Research on the classification and identification of uncertainty shocks, transmission mechanism and its impact on China's economic and financial stability [D]. Jilin University,2023.

[10] He Xiaodan, Chen Dingxi. Age structure and household financial allocation risk--an empirical study based on Chinese household financial survey data [J]. Journal of Nanjing University of Finance and Economics, 2020(03): 73-83.