ISSN 2616-5902 Vol. 6, Issue 5: 141-146, DOI: 10.25236/AJBM.2024.060519

Research on Vegetable Cost Pricing and Sales Forecast Based on PSO Optimization and Forecast Model

Ziting Zhu, Boyu Wu, Tian Wang

College of Economics and Management, Northeast Agricultural University, Harbin, 150006, China

Abstract: *Due to the short shelf life of vegetable items in fresh produce superstores, most varieties need to be restocked on a daily basis. In order to make reasonable pricing and replenishment decisions, this paper firstly integrates data and establishes a multiple linear regression model to find out the relationship between price and sales volume, and further optimizes it using PSO particle swarm algorithm to obtain the linear relationship equation, and then establishes an ARIMA time-series prediction model using SPSS to predict the total amount of replenishment and the pricing strategy for July 1-7 by using the data of the last 4 weeks. When considering the number of sales category constraints, this paper carries out the single-product profit ranking and selects the top 33 single-products with higher profits by combining the available varieties on the basis of meeting the minimum display quantity, and predicts the replenishment and pricing strategy on July 1 using the GWO gray prediction model. Ultimately, based on the above conclusions, the strategy for superstores to maximize returns in pricing and replenishment is derived.*

Keywords: *Multiple Linear Regression, PSO Particle Swarm Algorithm, Supermarket Sales*

1. Introduction

In fresh produce superstores, the freshness period of general vegetable commodities is relatively short, and the quality deteriorates with the increase of sales time, and most of the varieties cannot be re-sold on the next day if they are not sold on that day. Therefore, supermarkets usually restock and price their products daily based on historical sales and demand. Since supermarkets sell many different varieties of vegetables from different origins, and vegetables are usually purchased from 3:00 a.m. to 4:00 a.m., merchants are required to make replenishment decisions for each vegetable category on the same day without knowing the specific product and purchase price.

In order to make reasonable pricing and replenishment decisions, we reviewed and organized the relevant literature. The price of vegetables is mainly affected by temperature, moisture, and topography in the production chain; by the perishability of the vegetables themselves and transportation conditions in the distribution chain; and by holiday price increases and supply and demand in the sales chain[1];There is a close relationship between price and sales volume, which can be fitted to a linear equation with a strong negative linear correlation, i.e., when the price of oleaginous vegetables is high, the sales volume decreases, and when the price is low, the sales volume increases[2];The pricing of vegetables is generally based on the "cost-plus pricing" method. At the same time, the demand for fresh food is greatly affected by price and shelf allocation. As price-sensitive commodities, their pricing will directly affect retailers' profits and replenishment strategies, and the joint decision-making of pricing and shelf allocation can make more rational use of shelf resources, effectively reduce the losses caused by spoilage, and improve retailers' profits[3].Seasonal fluctuations in the prices of spicy vegetables are significant, with an overall upward trend in prices and cyclical fluctuations; the seasonal fluctuations in their prices are closely related to the growth cycle of the vegetables; random fluctuations do not show any significant growth or decline; there is a long-term and stable covariance between the prices of spicy vegetables, and the prices have a linkage, and the relationship is more complex[4].Supermarkets often sell discounted items for shipping damage and deterioration in quality. The relationship between the volume of merchandise sold and the time of day from a supply and demand perspective as well as the issue of the mix of merchandise sold are extremely influential on the superstore's selling. Reliable market analysis is particularly important for replenishment decisions and pricing decisions.

Data from http://www.mcm.edu.cn/index_cn.html.Based on this, in order to analyze the relationship between the total sales volume of various vegetable categories and cost-plus pricing and to formulate the total daily replenishment volume and pricing strategy of each vegetable category for the coming week, we first processed the merchandise sales data obtained for six vegetable categories distributed by a superstore and then established and optimized the multiple linear regression model; on the basis of which, we used the data of the last four weeks to predict the total daily replenishment volume and pricing strategy for the week of July 1-7 by using the ARIMA time-series model.

At the same time, due to the limited sales space of vegetable products, in order to further formulate the replenishment plan of single products, we set the total number of available single products to control at 27-33, and the order quantity of each item meets the minimum display amount of 2.5 kg, based on the profit ranking of the previous processed data,and combining with the data from June 24,2023 to June 30,2023, the top 33 items with high profit were selected.On the basis of meeting the minimum display volume and the total number of single products, the GWO prediction model is used to forecast the replenishment volume and pricing on July 1st, and on the premise of meeting the market demand for all kinds of vegetable products, the supermarket revenue is the maximum.

2. The Total daily replenishment and pricing strategy for each vegetable category for the coming week

2.1 Relationship between total sales and cost-plus pricing

Through the outlier and missing value processing on the commodity sales information data of 6 vegetable categories distributed by a supermarket,and integrating the attached data using the VLOOKUP function in excel, calculating the total sales.The data in the attachment are integrated with the VLOOKUP function in excel to calculate the total sales volume of each category and each single product, quarterly total sales volume, cost-plus pricing, and daily profit successively.And Quarterly sales and cost-plus pricing are then calculated by category. Then a multiple linear regression model is established, and the PSO particle swarm algorithm is used to further optimize the regression model[6]; finally, an ARIMA time series prediction model is established with the help of SPSS software on the basis of the regression linear equation, and the total daily replenishment and pricing strategy of the coming week are predicted by the data of the last 4 weeks.

Since the pricing of vegetables is generally based on the "cost-plus pricing" method, the data is first consolidated in an excel spreadsheet using the VLOOKUP function, and the total sales of each category and each individual product, the total quarterly sales, the cost-plus pricing, and the daily profit are calculated:

$$Cost - plus\ pricing\ =\ unit\ cost\ \left(1\ +\ cost\ margin\right) \tag{1}$$

$$unit\ cost = \frac{total\ cost}{sales\ volume} \tag{2}$$

$$\text{cost margin} = \frac{\text{price - wholesale price}}{\text{wholesale price}} \tag{3}$$

$$\text{cost-plus pricing} = \frac{\text{wholesale price } \times \text{sales} + \text{attrition rate} /100 \times \text{sales}}{\text{Sales Volume (1+ (price - wholesale price)/wholesale price)}} \tag{4}$$

Since there are six vegetable categories in the data given, it is too redundant to describe all of them directly, so quarterly sales and cost-plus pricing are calculated on a category-by-category basis, and then a multiple linear regression model is established[5].The PSO particle swarm algorithm is applied to further optimize the regression model on the basis of multiple linear regression to obtain the linear equation. It is shown as follows:

$$y_i = \beta_0 + \beta_1 x_{1cabbage} + \beta_2 x_{2\,philodendron} + \beta_3 x_{3\,Aquatic\ rhizomes} + \beta_4 x_{4eggplant} + \beta_5 x_{5capsicum} + \beta_6 x_{6edible\ fungi} \tag{5}$$

where,β_i is the coefficient of the independent variable and β_0 is the coefficient of the constant term.y_i denotes the total number of sales in the i^{th} quarter and x_i denotes the average cost plus pricing of each type of vegetable in the i^{th} quarter. We imported the collected data into MATLAB and used the self regress function to perform multiple linear regression to obtain the results:β_0 for 0,β_1 for 1.0023, β_2 for -0.4225,β_3 for -0.1077, β_4 for -0.7149, β_5 for 0.2020, β_6 for -0.7204;The p-test value of 0.032564 indicates that there is a linear correlation between total sales and cost-plus pricing;The R^2coefficient is 0.9997, indicating a very strong linear correlation. The final expression of the relationship is obtained:

$$y_i = 1.0023x_{1cabbage} - 0.4225x_{2philodendron} - 0.1077x_{3Aquatic\ rhizomes} - 0.7149x_{4eggplant} + 0.2020x_{5capsicum} - 0.7204x_{6edible\ fungi} \tag{6}$$

In order to more intuitively reflect the function expression, we use MATLAB to draw the fitting effect diagram, as shown in Figure 1:

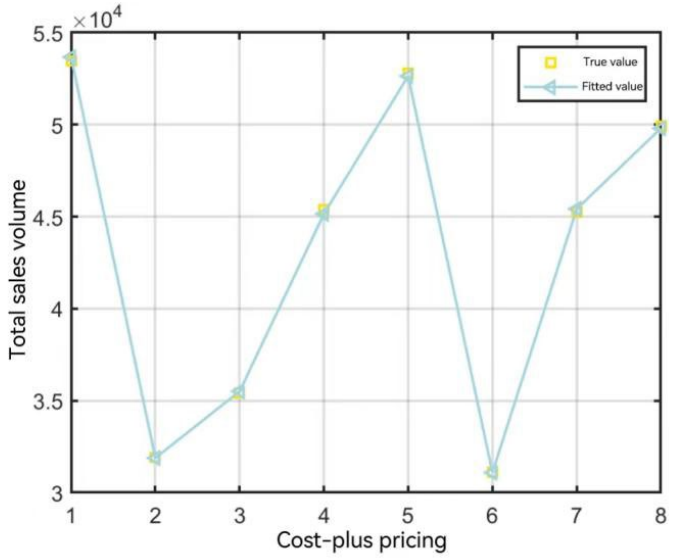

Figure 1: Fitting effect diagram

2.2 Total daily replenishment and pricing strategy

According to the consolidated data, consumption for the 30 days prior to July 1, 2023 is more stable and July 1-7, 2023 is a complete week. In order to ensure that the revenue of the superstore is maximized, then it is necessary to meet the sales demand of customers to avoid excess and insufficient sales space, i.e., too high a stocking volume leads to products over the freshness period and insufficient stocking volume leads to the inability to meet the demand of consumers. Therefore, this paper utilizes the sales data of the first 4 weeks (28 days) to directly predict the data of the last 7 days through a time series model.

For the first 4 weeks of the sales volume and daily average price of each vegetable category to produce time series charts and test the divergence smoothness, the autocorrelation plot (ACF) and partial autocorrelation plot (partial ACF) of the sequence of various types of vegetables most of the number is located inside the confidence interval, indicating that the sequence makes the basic smooth. So this paper uses SPSS software to establish ARIMA time series prediction model to dig the law of replenishment volume and pricing strategy in the first 4 weeks[7], and further predicts the data of the last 7 days directly, and the obtained results are shown in Table 1 and Table 2:

Table 1: Total daily replenishment

	philodendron	cabbage	Aquatic rhizomes	eggplant	capsicum	edible fungi
7.1	127.669	14.323	18.753	20.87	81.567	43.014
7.2	125.73	14.085	18.842	20.516	80.852	42.025
7.3	123.79	13.847	18.931	20.162	80.137	41.037
7.4	121.851	13.61	19.02	19.808	79.422	40.048
7.5	119.912	13.372	19.109	19.454	78.707	39.06
7.6	117.973	13.134	19.197	19.1	77.992	38.071
7.7	116.033	12.896	19.286	18.746	77.276	37.083

Table 2: Pricing Strategies

ISSN 2616-5902 Vol. 6, Issue 5: 141-146, DOI: 10.25236/AJBM.2024.060519

	philodendron	cabbage	Aquatic rhizomes	eggplant	capsicum	edible fungi
7.1	4.8833	13.0345	16.1594	7.3985	7.1525	16.389
7.2	4.846	12.9657	16.2227	7.3433	7.1564	15.838
7.3	4.8088	12.8968	16.2859	7.2881	7.1604	15.912
7.4	4.7716	12.8279	16.3492	7.2329	7.1644	17.757
7.5	4.7344	12.759	16.4124	7.1777	7.1683	19.028
7.6	4.6971	12.6901	16.4757	7.1225	7.1723	16.398
7.7	4.6599	12.6212	16.5389	7.0673	7.1762	15.1584

3. Pricing strategies and replenishment levels when superstores maximize revenue

3.1 The establishment of Gray Forecast Model

First assume a set of non-negative data series i.e. the original data

$$X^{(0)}=(x^{(0)}(1)\ ,\ x^{(0)}(2)\cdots x^{(0)}(\mathrm{n})\ ) \tag{7}$$

Introduce the weakening operator D such that

$$X^{(0)}D=(x^{(0)}(1)d,x^{(0)}(2)d\cdots x^{(0)}(n)d) \tag{8}$$

Including among those

$$x^{(0)}(k)d=\frac{1}{n-k+1}(x(k)+x(k+1)+\cdots+x(n)) \tag{9}$$

Then, $X^{(0)}DX$ the 1-AGO of

$$X^{(1)}D=(x^{(1)}(1),x^{(1)}(2)\cdots x^{(1)}(n))\qquad k=1,2,3\cdots n \tag{10}$$

Let $Z^{(1)}$be the sequence generated by the immediate neighborhood mean of $X^{(1)}$, then

$$z^{(1)}(k)=\frac{1}{2}(x^{(1)}(k)+x^{(1)}(k-1))\qquad k=2,3,4\cdots n \tag{11}$$

From there, the B and Y matrices can be constructed:

$$\mathrm{B}=\begin{bmatrix}-\mathrm{z}(2) & 1\\ \vdots & \vdots\\ -\mathrm{z}(\mathrm{n}) & 1\end{bmatrix}\qquad \mathrm{Y}=\begin{pmatrix}\mathrm{x}^{(0)} & (2)\\ \vdots & \vdots\\ \mathrm{x}^{(0)} & (\mathrm{n})\end{pmatrix} \tag{12}$$

Least squares estimation of the parameter u yields the estimates a (development coefficient), b (gray effect size), then the GM (1, 1) equation is given by

$$\frac{dx^{(1)}}{dt}+ax^{(1)}=b \tag{13}$$

This leads to the prediction model:

$$\hat{x}^{(1)}(k)=\left[x^{(0)}(1)-\frac{b}{a}\right]e^{-a(k-1)}+\frac{b}{a}\qquad k=1,2,3\cdots n \tag{14}$$

3.2 Pricing strategies and replenishment levels when superstores maximize revenue

Firstly, on the basis of data integration, the profit ranking of individual items is carried out, and the top 33 items with higher profits are selected on the basis of combining the available varieties from June

ISSN 2616-5902 Vol. 6, Issue 5: 141-146, DOI: 10.25236/AJBM.2024.060519

24, 2023 to June 30, 2023, and the replenishment and pricing strategies are predicted for July 1 using the GWO gray prediction model in 3.1 as shown in the table3 and pricing strategies are presented in Table 4:

Table 3: July 1st replenishment levels

Date	Ginger, Garlic & Millet Pepper Combo (Small portion)		Round Eggglant (2)	Sweet potato tip	Bisporus mushroom (box)	Yunnan oillettuce (portion)		Malabar spinach	Long-term eggplant	Cordyceps flower (part)
7.1	5.182		2.526	6.192	9.279	27.921		5.591	7.968	2.57
Date	Wild pink lotus root	Red pepper (2)	High melon (1)	Xixia Flower Mushroom (1)	Zhijiang green stem scattered flowers	High melon (2)	Net lotus root (1)	Rhomeric cabbage	Spinacia oleracea	Screw pepper
7.1	2.604	2.512	4.044	9.923	13.971	2.654	7.535	4.082	16.016	8.749
Date	broccol	White jade mushroom (bag)	Purple eggplant (1)	Red lotus root belt	Yunnan oil white vegetables	Screw pepper (part)	Yunnan Lettuce (portions)	Green and red hang pepper combination (parts)	Spinach (part)	Baby cabbage
7.1	18.593	2.69	2.523	2.927	27.921	13.377	38.32	3.159	16.016	12.584

Table 4: July 1st pricing strategy

Single name	pricing strategy (%)
Wild pink lotus root	166.39
Colorful pepper (2)	127.57
Red pepper (2)	130.15
High melon (1)	183.19
Xixia Flower Mushroom (1)	71.43
Zhijiang green stem scattered flowers	201.72
High melon (2)	85.95
Net lotus root (1)	77.78
Rhomeric cabbage	71.15
Spinacia oleracea	65.29
Screw pepper	69.01
Broccoli	51.68
Green eggplant (1)	268.10
White jade mushroom (bag)	108
Purple eggplant (1)	77.87
Red lotus root belt	73.58
Yunnan oil wheat vegetables	109.30
Screw pepper (part)	146.86
Yunnan Lettuce (portions)	143.70
Green and red hang pepper combination (parts)	139.67
Spinach (part)	208.27
Baby cabbage	64.65
Ginger, garlic, millet and pepper combination (small portion)	109.61
Yunnan lettuce	36.90
Round eggplant (2)	44.40
Sweet potato tip	66.67
Green line pepper (part)	72.66
Fresh fungus (part)	136.49
Bisporus mushroom (box)	52.78
Yunnan oillettuce (portion)	70.45
Malabar spinach	42.86
Long-term eggplant	16.28
Cordyceps flower (part)	68.14

Note: Here the pricing strategy applies only to (wholesale price * sales volume + attrition rate * sales volume) * (1 + pricing strategy)

ISSN 2616-5902 Vol. 6, Issue 5: 141-146, DOI: 10.25236/AJBM.2024.060519

4. Conclusions

In the case of considering the fresh business preservation period of vegetables, in order to formulate reasonable pricing and replenishment strategy. Undering the data integration in order to find out the relationship between price and sales, our paper establish multiple linear regression equation model.And then, using the PSO particle swarm algorithm further optimization, on the basis of the linear relationship using SPSS to set the time series prediction model, through nearly 4 weeks data forecast on July 1-7 replenishment amount and pricing strategy. Under the condition of the quantity of sales categories, this paper ranks the single product profit, thus obtains the replenishment quantity and pricing strategy on July 1, and finally obtains the pricing strategy of maximizing the single product profit ranking.Therefore, under revenue maximization, the July 1 pricing is ranked from high to low as Wild Lotus Root Powder, Colorful Peppers (2), Red Peppers, and Tall Melons, etc., and the replenishment is ranked from high to low as Yunnan Oilseed Vegetables, Yunnan Oilseed Vegetables (portions), and Broccoli.To sum up, this paper establishes PSO particle swarm algorithm and AMRIA time series model to help supermarkets establish relevant pricing strategies, provide reference for their pricing and replenishment, and have reference value for the pricing of other commodities.

References

[1] Mao Lisha. Research on the Pricing strategy and Production and marketing mode of vegetable wholesale Market from the perspective of supply chain [D].Central South University of Forestory and Technology, 2023.
[2] Yang Haoxu. Analysis of the correlation between vegetable price and sales volume—Take wheat cabbage as an example [J].The Food Guide, 2018(21): 179-181.
[3] Yang Shuai, Huang Xiangmeng, Wang Junbin. Research on the joint optimization strategy of shelf distribution and pricing of fresh food [J]. Supply chain management, 2022, 3(08): 49-59.
[4] Bai Li, Wang Mengwei, Zhou Yue, etc. Study on the price fluctuation characteristics of spicy vegetables in China [J].Northern gardening, 2023(24): 138-147.
[5] Cohen, J., Cohen P., West, S.G., & Aiken, L.S. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates. 2003.
[6] Li Xin. Chatic time series prediction based on particle swarm optimization algorithm [D].Dalian University of Technology, 2021.
[7] Yan Wenpeng, Wang Zhitao, Yuan Xiao. Time-series similarity measure based on fractional differentiation and its application[J].Journal of Sichuan University (Natural Science Editiom) ,2023, 60(04): 110-117