An Optimization Study of Pricing Strategies for Vegetable Category Based on Seasonal ARIMA Model with Linear Regression

Zhenshen Pan *

School of Management, Guangdong University of Science and Technology, Dongguan, China, 523083

* Corresponding Author Email: zhenshenpan@gmail.com

Abstract. Different categories of vegetables have different growth cycles, seasonal supply variations, and price fluctuations. Supermarkets need to consider these factors when making replenishment and pricing decisions in order to maximise revenue while meeting market demand. In this paper, replenishment and pricing decisions are made based on the historical sales data of the superstore, using a linear regression model and then predicting future sales based on a seasonal ARIMA model. A new mathematical model about pricing and an optimization model with the objective of maximising the revenue of the superstore are developed to get the pricing strategy mainly by adjusting the cost rate and give the total daily replenishment and pricing strategy for vegetables from 1 to 7 July 2023 for each category.

Keywords: Seasonal ARIMA model, Linear regression, Vegetable category pricing.

1. Introduction

Fresh food supermarkets sell a wide variety of vegetable varieties, and because of the qualities of vegetable commodities themselves - freshness period are shorter and commodity quality and sales time is inversely proportional to the [1-2], therefore, supermarkets often use the method based on the historical sales and demand for each commodity to carry out the daily replenishment [2-3], and often use the cost of pricing plus method for pricing vegetables [4]. It can be seen that the decision to replenish and price vegetables in supermarkets is related to whether supermarkets maximise their revenue or not. In order to develop a reasonable plan for replenishment and pricing of vegetable supermarkets, Yu Weige, Wu Huarui, Peng Cheng [4] constructed a combination model (L-BPNN) based on the combination of Lasso regression method and BP neural network based on influencing factors to predict the price of cucumber by comparing the Lasso regression model, BP neural network model, RBF neural network model and other regression analysis and intelligent analysis methods, and came up with the smallest average relative error, and was able to price the various categories of vegetables effectively; Liu Rui, Nie Ying, Sun Yongqi et al [5] systematically analysed and discussed the impact of various factors on the pricing of agricultural products from six aspects to provide a theoretical basis for reasonable pricing of agricultural products; Ye Jun, Gu Bojun, and Fu Yufang [6] obtained the existence of a unique optimal equilibrium strategy under FOB, CIF, and DAP trade modes by constructing the Stackelberg game model of fresh produce supply chain; Mao Lisha [7] and Lou Xinxin, Feng Aifen, Rong Bo [8] analysed how vegetables and other agricultural products can be scientifically priced from the macro-strategy by building an ARIMA model, which provides feasible ideas for the development of vegetable pricing strategy optimization.

Considering that superstores use categories as units for replenishment planning, analyse the relationship between the total sales of each vegetable category and the cost pricing mark-up method. First process to get the relevant sales information of each category of vegetables. Through the combination of sales unit price and total sales volume, the linear regression analysis based on the least squares method is carried out to analyse the sales volume of each vegetable category at each time of the day, to establish a linear regression model for a single category, and to test the model, based on which the linear regression model of each vegetable category is established. In this process, the relationship between the total sales volume of each category and cost-plus pricing can be known.

First of all, the flower and leaf category is analysed and the time series stability test is carried out, after obtaining the ADF test table to determine that the time series is a smooth time series, then the analysis is carried out in line with the demand for seasonal ARIMA model to solve, and after processing to obtain the model's parameter table, through which the total sales volume of each category of vegetables from 1-7 July 2023 is forecasted. The mathematical model on pricing is then developed to obtain the pricing strategy and the daily replenishment for the coming week. Finally, an optimization model is developed by analysing the findings and combining them with the mathematical model on pricing with the objective of maximising the revenue of the superstore.

2. Solving and analysing replenishment and pricing models

2.1. Model Preparation and Establishment

First use MATLAB to traverse the total sales table data of a particular superstore, data for a supermarket was obtained from the following web site: http://www.mcm.edu.cn/, to get the total sales volume of each vegetable category with the unit price of each item in the category of highest and lowest price.

Taking 1 July 2020 as an example, the analysis yields the total sales for each time period on that day as shown in Table 1.

Table 1. Analysis of sales by time period

time quantum	9	10	11	12	13	14	15
sales volume	27.146	58.213	46.262	19.317	13.897	10.221	13.897
time quantum	16	17	18	19	20	21	22
sales volume	19.337	47.512	33.718	49.542	42.444	29.236	7.501

By analysing the time slots, the basic sales slots are obtained, as shown in Figure 1, which demonstrates that the high sales slots are mainly in the part of the time slots of 10 o'clock and 17 to 20 o'clock.

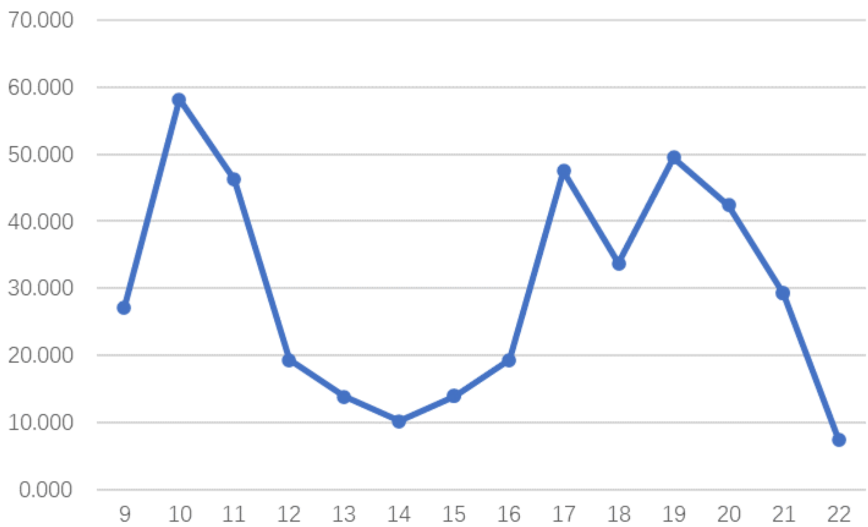

Figure 1. Total sales by time period

From the data analysis, it can be seen that only sweet potato tips were discounted on that day.Studying the sales of sweet potato tips from 1 July 2020 to 7 July 2020, it was found that since no sales were made on 4 July, sales rose sharply on 5 July and no discounting was done. This leads to the conclusion that surplus, and replenishment have a large impact on whether or not discounting occurs.

2.2. Modelling and solving

The results of the data processing of a superstore show that the unit sales price fluctuates according to the market. Get the average unit price:

$$J = \frac{\sum_{j=1}^{n} a_j}{n} \tag{1}$$

Where a_j is the unit price of the first j sale and n is the number of like codes.

The total category unit price was obtained by summing the average unit price of sales like for like Z, which was combined with the total number of sales of each vegetable category to perform a linear regression analysis based on the least squares method, with B being the coefficient in the case of having a constant.

$$standarderror = \frac{B}{t} \tag{2}$$

$beta$ represents the regression coefficient, the standardised regression coefficient represents the correlation between the independent variable, i.e. the predictor variable, and the dependent variable and allows for a comparison of the magnitude of the effect of the independent variable on the dependent variable in the case of a significant relationship. β value greater than 0 indicates a positive effect, and a larger value indicates a greater effect. β value of less than 0 indicates a negative effect, and a smaller value indicates a greater effect.

The beta derivation method is divided into regression method as well as formula method, the source of its formula method is the derivation of the regression method, so the formula method is consistent with the final result of the regression method, the formula of β is obtained through the information:

$$\beta = Cov(r_i, r_m)/(\sigma_m^2) \tag{3}$$

In summary, a linear regression model was developed on vegetables for each category:

$$y = \beta_0 + \beta_1 Z \tag{4}$$

Where β_0 takes the value of 21705.253 and β_1 takes the value of 123.141.

The formulae that summarise the linear regression model for leafy and flowering vegetable products are as follows:

$$y = 3116.354 - 121.485 * Z_1 \tag{5}$$

Combined with the other five categories of vegetables, the following linear regression model combining each category of vegetables was obtained:

$$\boldsymbol{y = \beta_0 + \beta_1 Z} \tag{6}$$

Where β is the regression coefficient and $\beta_0 = \begin{pmatrix} 3116.354 \\ 14356.788 \\ 1396.553 \\ 4951.743 \\ 2661.241 \\ 1245.529 \end{pmatrix}$, $\beta_1 = \begin{pmatrix} 121.485 \\ 454.81 \\ -60.443 \\ 293.771 \\ 57.169 \\ 14.266 \end{pmatrix}$.

As a result, the relationship between total sales and cost-plus pricing for each vegetable category tends to be linear.

In order to maximise the revenue of the superstore, the total daily replenishment should be equal to the predicted total sales of the next day, i.e. the total sales of each vegetable category from 1-8 July 2023 should be predicted, taking into account the condition "if it is not sold on the same day, it cannot be sold on the next day".

Extract and calculate the daily sales of each category for the month of June for analysis, starting with leafy and flowering vegetables. In seasonal ARIMA modelling, we need to first perform a smoothness test on the time series to ensure that it can be fitted by the ARIMA model.

If the time series is not smooth, we need to differentiate it or other treatments to make it smooth.ADF test is a unit root test to determine whether the time series is smooth or not by testing the existence or non-existence of the unit root of the series, so as to decide whether we need to differentiate it or not and other treatments. Calculations were made for each data and combined with SPSS analysis to produce Table 2.

Table 2. ADF test table

variant	sequences	t	P	AIC	threshold value		
					1%	5%	10%
philodendron	original sequence	-4.598	0.000***	201.07	-3.689	-2.972	-2.625
	1st order difference	-1.967	0.301	194.165	-3.77	-3.005	-2.643
	1st order difference - 1st order seasonal difference	-3.581	0.006***	180.093	-3.833	-3.031	-2.656
	2nd order difference	-7.789	0.000***	187.298	-3.77	-3.005	-2.643
	2nd order difference - 1st order seasonal difference	-3.793	0.003***	172.563	-3.859	-3.042	-2.661

Among them, the AIC, or Akaike Informativeness Criterion, is a measure of the goodness of fit of a statistical model, based on the concept of entropy, which allows weighing the complexity of the estimated model against the goodness of fit of this model to the data. In the general case, AIC can be expressed as:

$$AIC = \frac{(2k - 2L)}{m} \tag{7}$$

Where k is the number of parameters in the fitted model, L is the log-likelihood, and m is the number of observations.

The results of the test for this series showed that the original series based on flower and leaf categories showed a p-value of 0.000 at a significance level of 0.1%, level of significance, so the original hypothesis was rejected, and the series was a smooth time series. The data were subjected to an analytical solution that meets the demands of the seasonal ARIMA model and Table 3 was obtained.

Table 3. List of model parameters

term	ratio	standard error	z	P	95% lower confidence limit	95% upper confidence limit
ar. L1	0.366	0.342	1.073	0.283	-0.303	1.036
ar.S. L2	-0.884	0.325	-2.719	0.007***	-1.522	-0.247
ar.S. L4	-0.614	0.282	-2.18	0.029**	-1.167	-0.062
sigma2	1677.426	476.934	3.517	0.000***	742.653	2612.199

The above table shows the results of the model's parameters, including the model's coefficients, standard deviation, t-statistic results, etc., which can be used to analyse the model formula. The model parameter table predicts the eight data from 1 to 8 July, and the results of predicted sales for the next eight days are obtained as 114.851, 98.956, 127.677, 126.548, 128.733, 121.493, 119.913, and 109.008 (the prediction results are kept to 0.001).

Once the predicted data was obtained, Figure 2 was made to visualise the data and combine it with the original data that assisted in the prediction:

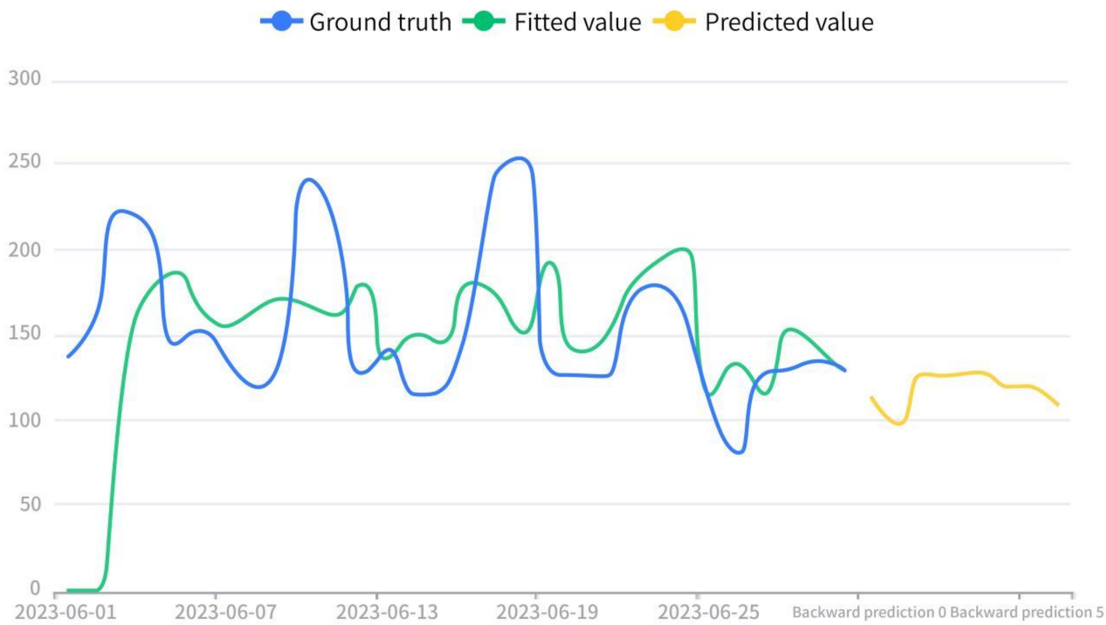

Figure 2: Forecasted sales for the next eight days

Pricing for the next seven days D is linked to the total daily replenishment R, the wholesale price G and the wastage rate S, so the pricing can be calculated using the following formula:

$$D = R(G + SG)(1 + \theta) \tag{8}$$

Where, θ is the cost interest rate, for the rest of the values, we can use the seasonal ARIMA model to speculate on the data, for the cost interest rate θ, from the title of the cost profit is equal to the profit divided by the unit price, combined with the above pricing formula can be deduced from the basic formula:

$$\theta = \frac{D - R(G + SG)}{D} \tag{9}$$

2.3. Testing of the model

Analysing the fit of the polynomial fit model using the coefficient of determination R^2 . R^2 and the adjusted R^2 is an elaboration of the model fit, and the adjusted R^2 is somewhat more accurate.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (\overline{y}_i - y_i)^2} \tag{10}$$

VIF is used to detect the severity of multicollinearity, can test whether the model presents covariance, the larger the value of VIF, the greater the possibility of covariance between independent variables, and its calculation formula is as follows:

$$VIF_i = \frac{1}{1 - R_i^2} \tag{11}$$

With the help of the following equations, the results obtained were calculated and the statistics of the results obtained were also introduced with the value of T .

T The value is the result of the t test of the regression coefficients, the larger the absolute value, the smaller the sig , sig represents the significance of the t test, and statistically, $sig < 0.05$ is generally considered to be a significant test of the coefficients. Significant means that the regression coefficients are significantly greater than 0 in absolute value, indicating that the independent variable can effectively predict the variance of the dependent variable.

The following is the formula for the degree of freedom ν where n is the number of samples and $\tau(x)$ is the gamma function, which can be obtained from the degree of freedom t :

$$\nu = n - 1 \tag{12}$$

$$t = \frac{\tau\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\tau\left(\frac{\nu}{2}\right)}\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{13}$$

From the past research [9], it is known that the interrelationship between each vegetable category is small, so the six categories can be analysed one by one, in this case the leafy vegetable category is used as an example for the linear regression analysis, and through the integration of the above information can be obtained in Table 4.

Table 4. Table of results of linear regression analysis

	Unstandardised coefficient		Standardised coefficient	t	P	VIF	R^2	Adjustment of R^2	F
	B	standard error	Beta						
a constant (math.)	21705.253	29793.622	-	0.729	0.507	-	0.597	0.497	F=5.934 P=0.072*
Total average sales price	123.141	50.551	0.773	2.436	0.072*	1			

In this case, from the analysis of the results of the F-test, it can be obtained that at a significance level of 1%, its p-value is 0.021, which presents significance at the level and rejects the original hypothesis that the regression coefficient is 0. Therefore, the model basically meets the requirements.

For variable covariance performance, all VIFs were less than 10, so the model had no multicollinearity problem, and the linear regression model for each category of vegetables was well constructed.

3. Results

3.1. The establishment of simulation model

Combining the above models, the returns are derived from the data of six different vegetable categories to build an optimization model with the objective of maximising the returns of the superstore:

$$\max\ z=\sum_{k=1}^{6} D_k R_k$$

$$\begin{cases} D_k=\dfrac{G_k(1+\theta_k)}{1-S_k} \qquad k=1,2,\cdots,6 \\ \theta_k=\dfrac{D_k(1-S_k)-G_k}{G_k} \qquad k=1,2,\cdots,6 \\ R_j=X_{j+1} \qquad j=1,2,\cdots,n \end{cases} \tag{14}$$

3.2. Analysis of experimental results

In summary, it can be seen that the pricing strategy is mainly through the adjustment of the cost rate of interest for pricing, the calculation of the rate of interest for each category of vegetables, combined with the daily replenishment volume of each vegetable category in the coming week, and consider the relevant literature [10] research on the characteristics of the price trend and the factors affecting the system to make the daily replenishment volume and the cost of interest rate of pricing on the day of the summary of the results in Table 5.

Table 5. 2023.7.1-2023.7.7 Replenishment Volume and Pricing Strategy Table

times	philodendron		cauliflower		Aquatic rhizomes	
	Total daily replenishment	pricing strategy	Total daily replenishment	pricing strategy	Total daily replenishment	pricing strategy
7/1	208.9	64.2%	42.5	31.3%	31.3	54.8%
7/2	189.6	75.5%	42.8	31.7%	28.4	61.6%
7/3	137.0	76.6%	29.2	32.5%	14.9	64.1%
7/4	129.7	76.0%	28.8	32.6%	14.5	69.0%
7/5	135.5	74.7%	30.3	31.2%	15.7	71.6%
7/6	123.2	87.6%	29.7	31.5%	15.1	78.8%
times	eggplant		chilli		edible fungi	
	Total daily replenishment	pricing strategy	Total daily replenishment	pricing strategy	Total daily replenishment	pricing strategy
7/1	29.2	117.2%	107.7	33.5%	88.4	78.9%
7/2	29.2	122.0%	100.7	46.6%	83.4	79.0%
7/3	22.2	129.3%	72.3	55.6%	58.1	83.0%
7/4	20.8	146.1%	68.6	71.7%	54.1	90.4%
7/5	20.8	145.0%	69.9	65.3%	60.7	97.2%
7/6	20.6	148.9%	69.4	73.5%	59.2	105.4%

In terms of replenishment volume and pricing strategy, the strategy is well constructed, and the pricing strategy model designed in this paper for the vegetable category can make the superstore maximise its revenue and fully satisfy the vegetable pricing demand.

4. Conclusions

Under the current market environment, vegetable commodities, as necessities, are often prone to loss of value due to changes in their own qualities.

Based on the fact that different categories of vegetables have different growth cycles, seasonal supply changes and price fluctuations, this paper conducts regression analysis on the historical sales data of supermarkets, establishes an optimal pricing model with the goal of maximising the revenue of supermarkets by combining with the seasonal ARIMA model, and provides new pricing strategies of vegetable categories for supermarkets in order to maximise the revenue.

References

[1] FANG Xin, YUAN Feng-jiao, CAI Ji-rong. An Optimized Coordination Model of Freshness-Keeping Efforts and Shelf Service for Fresh Produce Supply Chain[J]. Chinese Journal of Management Science, 2023, 31(6): 142-152.

[2] Zhang Yan, Mou Jinjin, Wang Shuyun. Optimization decision of supermarket fresh food supply chain with price control[J]. China Management Science, 2023, 31(10): 266-275.

[3] LIU Jianguo, WANG Junjin, ZHOU Jinxia, LIU Verhoeven. Research on supply chain contract coordination with different replenishment strategies under uncertainty environment[J]. China Management Science, 2019, 27(9): 68-79.

[4] YU Weige, WU Huarui, PENG Cheng. A combined model for short-term price prediction of vegetables based on Lasso regression and BP neural network[J]. Smart Agriculture, 2020, 2(3): 108.

[5] LIU Rui, NIE Ying, SUN Yongqi, SHENG Jiping, HUANG Jiazhang, WANG Pei, SUN Junmao. Factors Influencing the Pricing of High-quality Agricultural Products and the Pricing Strategy[J]. Journal of Agriculture, 2022, 12(3): 86-91.

[6] YE Jun, GU Bojun, FU Yufang. Cold chain logistics service and pricing decision of fresh produce supply chain under different trade modes[J]. China Management Science, 2023, 31(2): 95-107.

[7] Mao Lisha. Research on Pricing Strategy and Production and Marketing Mode of Vegetable Wholesale Market under the Perspective of Supply Chain [D]. Central South Forestry University of Science and Technology,2023.

[8] Lou Xinxin, Feng Aifen, Rong Bo. Research on marketing problems of agricultural products based on ARIMA model[J]. Modeling and Simulation, 2023, 12: 3918.

[9] Yao Yanmin, Bai Yuqi. Research on the framework of agricultural big data standard system[J]. Journal of Agricultural Big Data, 2020, 1(4): 76-85.

[10] YANG Juan, QIAN Tingting, ZHENG Xiuguo, et al. Characteristics of national and regional vegetable price trends and influencing factors[J]. Journal of China Agricultural University, 2021, 26(2): 188-198.