Optimization of vegetable commodity replenishment and sales strategy based on time series model and multiple linear regression model

Ke Xu[1, *], Wenqin Chen[2], Chengzhi Wang[2]

[1]College of Electronic and Information Engineering, Zhenjiang, 212100, China

[2]Jingjiang College of Jiangsu University, Zhenjiang, 212100, China

*Corresponding author: xkeyewitness@163.com

Abstract. Vegetable commodities have a short freshness period and their quality can change from good to bad. So the study of automatic pricing and replenishment decision of vegetable commodities is of great significance for the superstore itself. In this paper, we first discuss the total sales volume of each vegetable category (cost-plus pricing by single product selling price, wholesale selling price and wastage rate), and use the construction of multiple linear regression model to discuss the relationship between the total sales volume and the above three indicators. Next, the least squares estimation algorithm and Matlab were used to simulate and analyse the correlation between the cauliflower and foliage categories. In order to formulate the total daily replenishment and pricing strategy for the vegetable category in the coming week, a time series model was built to predict the total sales volume for the next 7 days by using 30 days of data changes. Finally, a moving average method is used to forecast the total daily replenishment amount, which maximises the supermarket's profitability.

Keywords: Vegetable commodities, multiple linear regression, time series forecasting.

1. Introduction

Vegetables are time-sensitive and expire quickly. Under this influence, the superstore will handle appropriate replenishment according to the daily sales and demand of these commodities [1-2]. In this paper, superstores do replenishment planning with vegetable categories, analyse the relationship between the total sales of different vegetable categories and cost-plus pricing, and make predictions about the daily replenishment and pricing measurements for each vegetable category from 1-7 July 2023 to maximise their superstore revenue.Data for this article comes from http://www.mcm.edu.cn/. The general idea of this paper: first of all, based on vegetable categories, analyse the relationship between the total sales volume of each vegetable category and cost-plus pricing [3-4], pricing is confirmed by the selling price of vegetable categories, wholesale price and attrition rate, take the first 10 weeks of vegetable single product data for regression model analysis, to get the data to be processed for relevance analysis, relevance analysis using the F-test of multiple linear regression equations, and then establish a Time series model, using one time moving average method, predicting the error value, and finally through the non-linear planning model, the conclusion is drawn.

2. Replenishment modelling and solving

Assume that the selling price, wholesale price and wastage rate of the independent variable category are a_{i1}, a_{i2}, a_{i3},the sales volume of the dependent variable category is and time is denoted by T_{i}. The dependent variable category is Y_1 and time is denoted by Y_2. The dependent variable category is Y_{i} and time is denoted by Y_3. And where Y_4 denotes cauliflower category, Y_5 denotes floral and leafy category, Y_6 denotes chilli category, a denotes eggplant category. [5-6].

For Y_1 cauliflower category, there are pairs of dataset $\{(Y_i, x_{i1}, x_{i2} \cdots, x_{ip}) : i = 1, \cdots, n\}$. If the following equation exists between the dependent and independent variables:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \tag{1}$$

Then the above equation is said to be the multiple linear regression model of the dependent variable Y_{i} on the independent variable $x_{1\ldots}x_p$, where β_0 is called the regression constant and $\beta_1 \cdots \beta_{\mathrm{p}}$ is called the regression coefficient

From the above analysis, it can be seen that the various categories are linearly correlated and the correlation is strong, so the regression model can be established as:

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \tag{2}$$

Computational analysis of the parameters: the least squares estimation is commonly used [7].

Calculations can be obtained by simulation analysis through *Matlab* software:

cabbage: β_0=330.967、β_1=-4.13369、β_2=-15.9497

philodendron: $\beta_0 = 1.7027$、$\beta_1 = -132.8333$、$\beta_2 = 42.7746$

So, we get the regression model:

$$\begin{aligned} Y_1 &= 330.967 - 4.13369x_{11} - 15.9497x_{12} + \beta_3 x_{13} + \varepsilon_1 \\ Y_2 &= 1.7027 - 132.8333x_{21} + 42.7746x_{22} + \beta_3 x_{23} + \varepsilon_2 \end{aligned} \tag{3}$$

Then,we first process the data for correlation analysis (under the condition that the significance level is set at 0.01 or 0.05, if the P value of the significance test of the correlation coefficient is less than 0.01 or 0.05, then it means that the explanatory variables are significantly related to the explained variables; if the P value is greater than 0.01 or 0.05, then it means that the explanatory variables do not have a significant correlation with the explanatory variables, and the corresponding explanatory variables should be excluded). (the corresponding explanatory variables should be eliminated). In a case-by-case analysis, the correlation between each independent variable and the dependent variable is analysed (an F-test can be performed). After eliminating those with poor correlation with the dependent variable, finally regression analysis is carried out. Relevant conclusions are obtained [8-9].

2.1. The F-test for multiple linear regression equations

The F-test for multiple linear regression equations focuses on testing the significance of the linear relationship between the independent variables and the dependent variable. Since the values of regression sum of squares and residual sum of squares vary with the sample size and the number of independent variables[10], the F statistic is constructed, namely:

$$F = \frac{SSR/p}{SSE/n-p-1} = \frac{MSR}{MSE} \sim F_\alpha(p, n-p-1) \tag{4}$$

Table 1. F-test results

ANOVA					
modelling	significance	degrees of freedom	mean square	F	significance
regression	5889.439	2	2944.719	4.987	.045
Residuals	4132.988	7	590.427		
Total	10022.427	9			

According to the analysis of Table 1, the regression sum of squares is 5889.439 and the residual sum of squares is 4132.988, the regression sum of squares plus the residual sum of squares equals to the total sum of squares of deviations is 10022.427, and the value of P is about 0.045, then it shows that the relationship between cauliflower category and pricing is significant.

The total daily replenishment and pricing strategy for each vegetable category for the coming week (1-7 July 2023) are given to maximise the return to the superstore. We choose the data of cauliflower category, and firstly, we carry out the data processing to get the total sales per day in June 2023, and build a time series model to predict the total sales in the next 7 days through the data change of 30 days, and use it as the data list of daily replenishment total.

2.2. Moving average method

For the time series model, we used one time moving average method for forecast calculation. Observation of the data for 30 days of data reveals significant variations that are not typical of N linear pattern. Therefore, our value of a is taken as 4(N)

Let the observed series be $y_1,\cdots,y_T$ and take the number of terms of the moving average $N < T$. The one time moving average is calculated as:

$$\begin{aligned} M_{\mathrm{t}}^{(1)} &= \frac{1}{N}\left(y_1 + y_{t-1} + \cdots + y_{t-N+1}\right) \\ &= \frac{1}{N}\left(y_{t-1} + \cdots + y_{t-N}\right) + \frac{1}{N}\left(y_t - y_{t-N}\right) = M_{t-1}^{(1)} + \frac{1}{N}\left(y_t - y_{t-N}\right) \end{aligned} \tag{5}$$

When the underlying trend of the forecasting target is to fluctuate up or down from a certain level, the forecasting model can be built using the one-shot moving average method:

$$\hat{y}_{t+1} = M_t^{(1)} = \frac{1}{N}\left(y_t + \cdots + y_{t-N+1}\right), t = N, N+1, \cdots, T, \tag{6}$$

Its standard error of prediction is:

$$S = \sqrt{\frac{\sum_{t=N+1}^{T}\left(\hat{y}_t - y_t\right)^2}{T-N}}, \tag{7}$$

The average of the sequence values for the most recent N period is used as a forecast for future periods. The value of N should be larger when the underlying trend of the historical series does not change much and when there is more random variation in the series, and vice versa. An effective way to choose the best N value is to compare the forecast errors of several models, the smaller the value of the forecast error, the better. The data for the primary mean shift analysis are shown in table 2.

Table 2. Primary mean shift analysis data

times	y_T	$M_t^{(1)}$	$\hat{y}_{t+1}$
1	19.364		
2	16.172		
3	33.243		
4	25.117	23.474	
5	22.57	24.2755	23.474
6	21.351	25.57025	24.2755
7	11.835	20.21825	25.57025
8	15.446	17.8005	20.21825
9	13.684	15.579	17.8005
10	26.766	16.93275	15.579
11	25.472	20.342	16.93275
12	17.143	20.76625	20.342
13	10.189	19.8925	20.76625
14	9.175	15.49475	19.8925
15	9.804	11.57775	15.49475
16	22.692	12.965	11.57775
17	26.791	17.1155	12.965
18	23.098	20.59625	17.1155
19	12.363	21.236	20.59625
20	13.919	19.04275	21.236
21	16.697	16.51925	19.04275
22	17.912	15.22275	16.51925
23	9.279	14.45175	15.22275
24	11.548	13.859	14.45175
25	10.681	12.355	13.859
26	8.083	9.89775	12.355
27	14.272	11.146	9.89775
28	16.069	12.27625	11.146
29	24.367	15.69775	12.27625
30	28.087	20.69875	15.69775
31			20.69875

The forecasting process of the total number of sales of cauliflower category on 1 July 2023 was carried out by *Matlab* software and its total sales were obtained as 20.69875.However, we need the total daily replenishment for the coming week (1-7 July 2023). Therefore, we made 7 forecasts by one time moving average method to get the data related to the next 7 days as shown in Table 3 below:

Table 3. Total sales of cauliflower category, 1-7 July

Time/Day	1 July	2 July	3 July	4 July	5 July	6 July	7 July
Total cauliflower sales/kg	20.69875	14.9547	12.6106	12.6161	13.7357	15.2568	16.4455

2.3. Data analysis

From the table 3, it can be found that by using this model to get the total amount of cauliflower category sales for the next 7 days, this is used to equate the total amount of cauliflower category sales for the 7 days to the total amount of daily replenishment for the coming week (1-7 July 2023). By looking at the data from 1 June - 30 June 2023 it is observed that the total cauliflower category sales fluctuate greatly from day to day and the data obtained from the forecast satisfies this, therefore this replenishment total can be used with this model.

The pricing strategy problem for the coming week (1-7 July 2023) is to find the profit of making the maximum sales per day for the next 7 days. As with the replenishment plan above, we have chosen to analyse only the data relating to each day in June 2023 for the cauliflower category. From these data metrics, it can be seen that profit = (1 - attrition rate) * total sales * disappearing unit price - total sales * wholesale price, where attrition rate and unit price of sales can be considered fixed values, total sales take a range by the values predicted above for the next 7 days, and wholesale price takes a range by the values for the previous 30 days. Thus, the wastage rate is equal to 14.142 per cent and the unit price is equal to 11.8918. Assuming that the profit is F, the total sales volume is X_1 belonging to (12.6106,20.6988) and the wholesale price is X_2 belonging to (8.083,33.243).

Thus, a non-linear planning model is obtained:

$$\max F = 10.2101 X_1 - X_1 * X_2$$

$$s.t. \begin{cases} 12.6106 \le X_1 \le 20.6988 \\ 8.0830 \le X_2 \le 33.243 \end{cases} \tag{8}$$

It is derived from *Matlab* software that when $X_1 = 13.6006, X_2 = 9.0730$ and when $F = 36.0047$,the quantity of cauliflower category stocked per day for the coming week can be taken to be about 13.6006 kg, the wholesale price is 9.0730 Yuan/kg, and the profit earned can be 36.0047 Yuan. This is used to speculate on the pricing decisions for other categories.

3. Conclusion

This paper discusses the replenishment strategies of superstores in dealing with vegetables with short freshness and variable quality. The study analysed the relationship between total sales volume and cost-plus pricing in terms of vegetable categories, and forecasted the daily replenishment volume and pricing strategy from 1 to 7 July 2023 to maximise superstore revenue. The selling characteristics of each vegetable category were examined through multiple linear regression and time series modelling, regression models were developed, and correlations and regression equations were tested.

Firstly, the regression relationship between the sales volume of each vegetable category was analysed using the selling price, wholesale price and wastage rate as independent variables. The calculation of the regression model was analysed using the cauliflower category as an example. The linear correlation of each category was verified through correlation analysis and regression equation test. Then, a time series model was developed using the one-time moving average method to predict the total sales volume for the next 7 days, which was used to develop a replenishment plan.

Finally, the optimal total daily replenishment and pricing strategy were derived from the nonlinear programming model. Taking the cauliflower category as an example, based on profit maximisation as the objective, a non-linear planning model was established to infer the daily replenishment quantity and optimal wholesale price for the coming week based on the range of predicted total sales quantity and the range of values of wholesale price. According to the model, it was inferred that the incoming quantity of cauliflower category in the coming week could be around 13.6006 kg, and the wholesale price would be RMB 9.0730/kg, resulting in a profit of RMB 36.0047. This study, through data modelling and analysis, provides recommendations on the amount of vegetables to be restocked and the optimal pricing for the coming week by the superstores in order to achieve the maximum profitability.

References

[1] Cai Jianhu,Cao Weizang,Zhou Qian et al. Research on VMI inventory decision-making model for assembly system under supplier output uncertainty environment [J/OL].China Management Science, 1-14[2023-11-08]

[2] Tianjin's "food basket" stable in autumn [N].Tianjin Daily, 2023-08-24 (010).

[3] Zheng L.Research on dynamic bulk inventory replenishment problem of community group-buying retailers [D].Zhongnan University of Economics and Law, 2022.

[4] Lai ZK.Beijing Fengtai makes every effort to protect residents' "food basket"[N].Workers' Daily, 2022-04-26 (004).

[5] Wu YAN.Increasing replenishment frequency to ensure personnel input [N].Shenzhen Special Zone News, 2022-03-15 (A03).

[6] Liang, Shuting.Research on the inventory strategy of community fresh food retail terminal where time and stock quantity jointly affect spoilage rate [D].Beijing Jiaotong University, 2020.

[7] Qiao Yannan.Research on the optimisation scheme of Q supermarket outlet layout based on improved SLP [D].Tianjin University of Finance and Economics, 2020.

[8] Epidemic now, supermarket fresh C position "leading the way", these fresh basic skills must be strengthened [J].Corporate Observer, 2020, (03): 111-113.

[9] ZHANG Jingfang,XIANG Wanli.Application of dynamic buffer inventory model in inventory management of agricultural products [J].Logistics Technology, 2018, 37 (12): 83-88.

[10] HAN HAN GWANG.Research on inventory management programme of frozen products category in community vegetable direct shops in Urumqi [D].Xinjiang Agricultural University, 2017.