A Study on Replenishment Decision of Vegetable Commodities Based on Time Series Forecasting and Linear Programming Models

Botao Jing[1, #], Huisong Jiang[2, #], Ziyi Wei[3, #, *]

[1]School of Information and Communication Engineering, Dalian University of Technology, Dalian, China, 116024

[2]School of economics and management, Dalian University of Technology, Dalian, China, 116024

[3]Business School, Dalian University of Technology, Dalian, China, 116024

* Corresponding Author Email: 19177248170@163.com

[#]These authors contributed equally.

Abstract. The automatic pricing and replenishment decisions for vegetable products are based on the fact that vegetable products have relatively short shelf lives, and there is a wide variety of vegetable types for sale. Vegetable purchases typically occur in the early morning. Therefore, businesses often rely on historical sales and demand data for each product to make pricing and restocking decisions, even when they do not have exact information about individual item and purchase prices. This article is based on the data provided by the CUMCM to conduct relevant research.

Keywords: Linear Programming, Time Series Forecasting Replenishment Decision.

1. Introduction

By analyzing daily procurement data, sales figures, and profit margins, we have developed a linear system of equations to model the relationship between total sales volume and cost-based pricing. This approach has led to the creation of an optimal replenishment and pricing model, aimed at maximizing the profit of the supermarket. Within this model, various constraints are considered, such as controlling the total number of individual products available for sale within the range of 27 to 33, and ensuring that the order quantities for each individual product meet the minimum display requirement of 2.5 kilograms.

Furthermore, through the adjustment of certain parameters, we have refined the model to better align with real-world conditions. The revised model takes into account the dynamic nature of sales, procurement, and customer demand, ultimately enhancing its accuracy and practicality.

2. Replenishment totals and pricing strategies for the vegetable category

The autoregressive integer moving average, ARIMA [1], is a statistical analysis model that uses time series data to forecast future trends. Over time, the sales volume of each category and the wholesale price for the coming week, formed by the forecast, are considered as a random sequence, and the model is used to approximate the description of the sequence. Once this sequence is determined, ARIMA predicts the future values based on the past and present values of the time series. The sales volume and wholesale prices of each category from July 1, 2020 to June 30, 2023 are cleaned from the data and the outliers are removed using ARIMA model to predict the sales volume as well as wholesale price of each category for the coming week. The ARIMA model consists of an autoregressive model and a moving average (MAA) model. The AR model describes the relationship between the current value and the lagged value, and predicts the future value using historical data. The MA model utilizes a linear combination of past residual terms to look at future residuals.

The ARIMA [2]prediction models can be written as the following equation:

$$\hat{p}^{\{t\}} = p_0 + \sum_{j=1}^{p} \gamma_j \, p^{\{t-j\}} + \sum_{j=1}^{q} \vartheta_j \varepsilon^{\{t-j\}} \tag{1}$$

For the price time series $\{x1, x2, - - - , xn\}$ With the help of **ARIMA** model, the lattice strategy is optimized by simply predicting the next day's price using past data in order to move the lattice according to the prediction given by **ARIMA**. The movement of the grid is determined by the weighted sum of the long term metrics of MA and the short term metrics in the model. The movement of the grid is calculated by the following equation:

$$ACF(q) = \frac{Cov(X_j, X_{j-p})}{Var(X_0)} = \frac{\frac{1}{n-p}\sum_{j=q+1}^{n}(x_j - \bar{x})(x_{j-p} - \bar{x})}{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2} \tag{2}$$

$$\widetilde{p_i}^{\{t\}} = p_i + \omega\left(MA^{\{t\}}(N) - MA^{\{t-1\}}(N)\right) + \mu\left(ARIMA^{\{t+1\}} + ARIMA^{\{t\}}\right) \tag{3}$$

The parameters and control the weights of the two metrics during the lattice movement. These two parameters will be adaptively adjusted during the backtesting phase by initializing to 0.3 and to 0.3 in the first cycle. Therefore, the final made by the adjusted lattice model to predict the weekly sales volume of each category as well as the wholesale price, the results are shown in Table 1.

Table 1. ARIMA model prediction results

(a) Future - - Saturday large category prices				(b) Future - - Large category sales on Saturday			
dates	cauliflower (Brassica oleracea var. botrytis)	philodendron	chilli	dates	aquatic stem	philodendron	cauliflower (Brassica oleracea var. botrytis)
20230701	7	11	8	20230701	19	190	31
20230702	7	11	8	20230702	16	188	31
20230703	7	11	8	20230703	14	182	28
20230704	7	11	8	20230704	13	177	28
20230705	7	10	8	20230705	16	176	26
20230706	7	10	8	20230706	19	181	27
20230707	7	10	8	20230707	20	187	28

denotes the cost-plus price, where = 1, 2, 3, 4, 5, 6 denotes the six categories of cauliflower, foliage, pepper, eggplant, tomato, edible fungi, and aquatic roots and tubers, respectively, and ,1 and ,1 denote the fit coefficients of the 1st and the 0th terms of the categories of ,respectively. It is assumed that there is a linear relationship [3] between the cost-plus price of category and the total sales of :

$$y_i = a_{i1}x + a_{i0}, i = 1,2,3,4,5,6 \tag{4}$$

$$y_{i1}x_i + a_{i0} = y_{i,j} i = 1,2,3,4,5,6 \tag{5}$$

The fitting objective is to minimize the sum of squares of the residuals, is:

$$Minimize \sum_{i=1}^{n}\sum_{j=1}^{6}\left(y_{i,j} - \left(a_{j,1}x_i + a_{j,0}\right)\right)^2 \tag{6}$$

To solve for the coefficients xi and yi [3]the regular equations using the least squares method are solved as follows:

$$y_i = a_{i,1}x + a_{i,0} i = 1,2,3,4,5,6 \tag{7}$$

where X is the design matrix and Y is the response matrix corresponding to the function [4]. The fitted function is solved as shown in Table 2

The following calculates ($MeanRelativ\ eError$) to test the effectiveness of the model fit, collect the actual data points (xi, yi) [5][6]and then use the fitted equation to
to calculate the predicted value $y\hat{}i$: :

$$\hat{y}_i = a_{i1}x_i + a_{i0} i = 1,2,3,4,5,6 \tag{8}$$

Calculate the relative error for each data point ($RelativeErr\ or$): (9)

$$RE_i = \frac{|y_i - \hat{y}_i|}{y_i} i = 1,2,3,4,5,6 \tag{9}$$

Calculate MRE ($MeanRelativ\ eError$):

$$MRE = \frac{1}{6}\sum_{i=1}^{6} RE_i \tag{10}$$

Solve for as shown in Table 2:

Table 2. Fitted equations and MRE data for different categories

kind	(math.) simultaneous equations	MRE
edible mushroom	y= -0.0096x + 316.9042	16
eggplant	y= -0.1190x+ 377.4092	20
chilli	y= -0.0486x + 533.0039	27
cauliflower (Brassica oleracea var. botrytis)	y= -0.0492x + 358.6750	16
philodendron	y= -0.0121x+ 312.9688	20
Aquatic rhizomes	y= -0.01258x + 544.1672	24

The cost-plus pricing method is a pricing strategy that, although simple and limited, allows for rapid decision-making in vegetable markets where the degree of market variability is relatively low

$$y_{i,j} = C_{ij}\left(1 + r_{ij}\right) \tag{11}$$

is used to calculate the pricing (usually the selling price) of a product in category at a given date. This price is calculated based on the unit cost of the product, , and the markup rate, .

$$C_{ij} = \frac{F_{ij} + V_{ij}}{Q_{ij}} \tag{12}$$

Calculate the unit cost of product for category on day. unit cost is equal to fixed cost and variable cost divided by sales volume.

$$r_{ij} = \frac{S_{ij} - W_{ij}}{W_{ij}} \tag{13}$$

Calculate the markup rate for products in category on day. The markup rate represents the difference between the selling price and the wholesale cost to the wholesale cost ratio. The markup rate helps to assess the difference between the selling price and the cost. Then the following relationship is obtained based on cost pricing:

$$y_{ij} = \frac{\left(W_{ij}Q_{ij} + y_{ij}\left(1 - L_{ij}\right)\right)}{x_{ij}}\left(1 + r_{ij}\right) \tag{14}$$

The profit maximization is expressed as follows [7]:

$$\max_{p_{ij}}\left(q_{ij}\left(p_{ij} - w_{ij}\right) - \left[W_{ij}Q_{ij} + P_{ij}\left(1 - L_{ij}\right)\right]\right) \tag{15}$$

Constructing constraints in the face of the complexity of the supply chain, and character shipment losses, preserving high inventory turnover and other realities require companies to quickly set prices

and replenishment decisions under uncertainty, determine markups based on the expected sales volume to make decisions quickly, and at the same time in the vegetable market, where demand fluctuations are relatively low to maintain a good market adaptability and competitiveness, and maximize the market demand, reduce losses, and Increase profitability. The criteria for determining the markup rate based on expected sales volume are as follows:

Type 1 expected sales volume, qij < 15 kg, markup rate $155 \leq rij1 \leq 165$ (most of the superstores are low-market, high-profit goods with higher profit per unit of goods)

2nd expected sales volume, qij < 45 kg, markup rate $135 \leq rij2 \leq 145$ (medium-volume merchandise that guarantees sales while earning more profit)

3rd Expected sales volume, $qij \geq 45$ kg, markup rate $115 \leq rij3 \leq 125$ (high volume goods, thin margins through lower prices) Based on the linear relationship, the following constraint is constructed:

$$y_i = a_{i0} + a_{i1}x_i i = 1,2,3,4,5,6 \tag{16}$$

The planning model is as follows: [8]

$$\max_{q_{ij}}\left(q_{ij}\left(p_{ij} - w_{ij}\right) - \left[W_{ij}Q_{ij} + P_{ij}\left(1 - L_{ij}\right)\right]\right) \tag{17}$$

$$s.t.\begin{cases} y_{ij} = \dfrac{\left(W_{ij}Q_{ij} + y_{ij}\left(1 - L_{ij}\right)\right)}{x_{ij}}\left(1 + r_{ij}\right)x_{ij} \\ y_{ij} = \left(1 + r_{ijk}\right)c_{ij} \\ x_{ij}, \quad w_{ij} \geq 0 \\ 155\% \leq r_{ij1} \leq 165\% \\ 135\% \leq r_{ij2} \leq 145\% \\ 115\% \leq r_{ij3} \leq 125\% \\ j = 1,2\cdots,7 \\ i = 1,2\cdots,6 \end{cases} \tag{18}$$

The total daily replenishment and pricing strategy for the six categories for the coming week are solved as shown in the table 3:

Table 3. Total daily replenishment and pricing strategy for the six categories for the coming week

dates	cauliflower (Brassica oleracea var. botrytis)		philodendron		chilli	
	Total daily replenishment	pricing strategy (%)	Total daily replenishment	pricing strategy	Total daily replenishment	pricing strategy (%)
2023-07-01	56	53	223	31	76	78
2023-07-02	20	24	251619	27	112	106
2023-07-03	43	48	272	41	109	108
2023-07-04	39	89	156	19	44	45
2023-07-05	51	43	272	88	110	113
2023-07-06	33	42	260	38	107	106
2023-07-07	28	55	199019	58	70	72
dates	eggplant		edible mushroom		Aquatic rhizomes	
	Total daily replenishment	pricing strategy(%)	Total daily replenishment	pricing strategy	Total daily replenishment	pricing strategy(%)
2023-07-01	10	93	67	32	23	83
2023-07-02	9	39	70	96	11	47
2023-07-03	5544	62	30	103	16	86
2023-07-04	10	101	52	83	10	29
2023 -07-05	3	100	44	88	19	56
2023-07-06	6965	104	60	88	10	34
2023-07-07	6	80	59	57	15	41

Suppose there are vegetable commodities, each with two variables: replenishment quantity,revenue, while the following conditions need to be satisfied:

(1) The total number of individual items is kept between 27 - 33.

(2) The minimum display quantity of 2.5 kg for each item ordered is met.

Mathematical modeling objective function: maximize total return

$$max \sum_i (\beta_i - \eta_i) x_i \tag{19}$$

Constraints.

(1) The total number of individual items is controlled to be between 27 - 33.[9]

(2) The order quantity of each individual item meets the minimum display quantity of 2.5 kg.

Range constraints on pricing strategy:

The simplex method of solving transforms a linear programming model into standard form: it ensures that the objective function is a maximization problem and that all constraints are equations [10] .

$$Maximize\ Z = \sum (\beta_i - \eta_i) z_j$$
$$s.t. \begin{cases} 27 \le \sum z_i \le 33 \\ z_i \ge 2.5 \ \ \forall i \\ p_i \le 100\% \quad \forall i \end{cases} \tag{20}$$

Solving the simplex table .Initialize the basic feasible solution(BFS):

(1) select the initial solution such that all constraints are satisfied. Usually, the initial solution can be chosen to be the zero vector.

(2) Select the daily average data 26 of the row where 102900011033944 with high cost-plus rate is located as the initial solution.

$$\begin{cases} max \sum_i (\beta_i - \eta_i) x_{i_{average}} \\ p_{average} \\ \beta_{average} \\ \eta_{average} \end{cases} \tag{21}$$

(3)Calculate the simplex table, Create the simplex table, including the rows of the objective function, the rows of the constraints, and the initial form of the table of relaxation variables as follows.

$$Z = -\eta_B + \sum \eta_j z_j \tag{22}$$

$$z_B = B - \sum A_i x_j \tag{23}$$

where is the coefficient of the objective function of the basic variable, is the coefficient of the objective function of the non-basic variable, is the value of the basic variable, is the value of the the values of the right-hand side of the constraints, and is a column of the coefficient matrix.

(4) Checking the optimality condition: check if there are non-negative jB in the rows of the objective function. If all ηj - ηB≤ 0, then the current solution is optimal, otherwise continue.

(5) Selection of base variables: Select a non-base variable such that - > 0. Usually, select a non-base variable such that - is maximized as the in-base variable.

(6) Selection of the outgoing base variable: A base variable is selected as the outgoing base variable in order to keep the new base solution feasible. This usually involves choosing the smallest /-ratio to ensure that the solution in the next step still satisfies the constraints.

(7) Update the Simplex Method table: Update the values in the table according to the selected in-base and out-base variables.

(8) Repeat Steps 4 - 7: Repeat Steps 4 through 7 until there are no positive - values in the rows of the objective function, at which point the optimal solution is reached.

(9) Obtaining the optimal solution: once the optimal solution is reached, the optimal solution is obtained by reading the values of the basic variables and the values of the non-basic variables in the table.

(10) Calculate the optimal objective function value: calculate the optimal objective function value, which is the maximized total return. By programming , the final

the optimal solution is obtained as in Table 4 :

Table 4. Table of daily replenishment and pricing strategy for individual products

product code	Total daily replenishment (KG)	Pricing strategy (%)	product code	Total daily replenishment (KG)	Pricing strategy (%)
102900011034330	10	53	102900011034439	5	59
102900005115779	10	134	102900011035740	1	86
102900005115878	9	142	102900011013274	2	125
102900011021842	7	71	102900011023464	5	128
102900005115861	18	94	102900011030059	13	98
102900005116530	18	47	102900011030097	9	94
102900011030905	1	63	102900011030110	15	138
102900005116899	25	128	102900011008164	6	36

3. Conclusion

This study successfully integrates time series forecasting and linear programming to optimize vegetable commodity replenishment and pricing strategies. The use of ARIMA models for predicting sales trends and pricing strategies in response to market variability has proven effective. Our model addresses the challenges unique to the perishable nature of vegetables and diverse product range, ensuring operational efficiency and profitability. The findings emphasize the need for rapid, adaptable decision-making in retail management, offering valuable insights for supermarkets and extending to other perishable goods markets. This research provides a significant contribution to enhancing efficiency in the dynamic retail sector.

References

[1] Xu J,Lin XY. Research on the implementation effect of fiscal and tax policy tools in China's sports industry--analysis based on time series data from 2008-2018[J/OL]. Journal of Shandong Institute of Physical Education:1-9[2023-10-31].

[2] WEN Jing,ZHANG Peng,CAO Yingnan et al. Study on the trend of hospitalisation cost changes of hip and knee arthroplasty patients based on interrupted time series analysis[J]. China Hospital Management,2023,43(10):42-47.

[3] TAO Xiaozhao,LI Jiapeng,YANG Mengzhe et al. Study on growth performance and growth curve fitting of F1 generation of otter and Igloo meat rabbit crosses[J/OL]. Chinese Journal of Animal Husbandry.

[4] Li P,Liu Yanna. Construction of hydraulic turbine operating characteristic surface based on moving least squares[J/OL]. Hydropower Energy Science,2023(11):183-186+17[2023-10-31]

[5] An Minjun. Application of Python based least squares method in electronic balance[J]. Branding and Standardisation,2023(06):175-177.

[6] WANG Peng,XIN Peikang,LIU Yin et al. Node centre coordinate extraction of mesh shell structure point cloud nodes using the least squares method[J/OL]. Journal of Graphics:1-9[2023-10-31].

[7] YANG Leun,LI Jingru,JIA Zhikun et al. Research on multi-objective path planning method for mobile operating robots based on operability[J/OL]. Control and Decision Making:1-10[2023-10-31]

[8] Xiaoying Wang,Jing Xu,Zhaoyu Yang et al. Multi-objective planning optimisation of isolated island integrated energy system considering carbon emission[J]. Thermal Power Generation,2023,52(10).

[9] Lei Hang, Liu Hesheng, Zhang Ruigang et al. Research and simulation of independent pitch control technology for wind turbines based on simplex method[J]. Thermal Power Generation,2023,52(03)

[10] TANG Dingyi. Facility siting optimisation algorithm in communication networks based on simulated annealing and network simplex method[J]. Computer Applications and Software,2022,39(08):125-131.