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

Botao Jing¹, #, Huisong Jiang², #, Ziyi Wei³, #,*

¹School of Information and Communication Engineering, Dalian University of Technology, Dalian, China, 116024
²School of economics and management, Dalian University of Technology, Dalian, China, 116024
³Business 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:
\[
\hat{p}^{(t)} = p_0 + \sum_{j=1}^{p} \gamma_j p^{(t-j)} + \sum_{j=1}^{q} \theta_j \varepsilon^{(t-j)}
\]  

(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:

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

(2)

\[
\hat{p}_i^{(t)} = p_i + \omega \left( MA^{(t)}(N) - MA^{(t-1)}(N) \right) + \mu \left( ARIMA^{(t+1)} + ARIMA^{(t)} \right)
\]

(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

<table>
<thead>
<tr>
<th>dates</th>
<th>cauliflower (Brassica oleracea var. botrytis)</th>
<th>philodendron</th>
<th>chilli</th>
<th>dates</th>
<th>aquatic stem</th>
<th>philodendron</th>
<th>cauliflower (Brassica oleracea var. botrytis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20230701</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>20230701</td>
<td>19</td>
<td>190</td>
<td>31</td>
</tr>
<tr>
<td>20230702</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>20230702</td>
<td>16</td>
<td>188</td>
<td>31</td>
</tr>
<tr>
<td>20230703</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>20230703</td>
<td>14</td>
<td>182</td>
<td>28</td>
</tr>
<tr>
<td>20230704</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>20230704</td>
<td>13</td>
<td>177</td>
<td>28</td>
</tr>
<tr>
<td>20230705</td>
<td>7</td>
<td>10</td>
<td>8</td>
<td>20230705</td>
<td>16</td>
<td>176</td>
<td>26</td>
</tr>
<tr>
<td>20230706</td>
<td>7</td>
<td>10</td>
<td>8</td>
<td>20230706</td>
<td>19</td>
<td>181</td>
<td>27</td>
</tr>
<tr>
<td>20230707</td>
<td>7</td>
<td>10</td>
<td>8</td>
<td>20230707</td>
<td>20</td>
<td>187</td>
<td>28</td>
</tr>
</tbody>
</table>

denotes the cost-plus price, where \(i = 1, 2, 3, 4, 5, 6\) denotes the six categories of cauliflower, foliage, pepper, eggplant, tomato, edible fungi, and aquatic roots and tubers, respectively, and \(a_{10}, a_{01}\) and \(a_{10}, a_{01}\) denote the fit coefficients of the 1st and the 0th terms of the categories of \(i\), respectively. It is assumed that there is a linear relationship [3] between the cost-plus price of category and the total sales of:

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

(4)

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

(5)

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

\[
\text{Minimize} \sum_{i=1}^{n} \sum_{j=1}^{6} (y_{i,j} - (a_{j1} x_i + a_{j0}))^2
\]

(6)

To solve for the coefficients \(x_i\) and \(y\) [3]the regular equations using the least squares method are solved as follows:

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

(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 \(\text{MeanRelative Error}\) to test the effectiveness of the model fit, collect the actual data points \((x_i, y_i)\) \([5][6]\) and then use the fitted equation to calculate the predicted value \(y_i\) :

\[
y_i = a_{i1}x_i + a_{i0}i = 1,2,3,4,5,6
\]

(8)

Calculate the relative error for each data point \(\text{Relative Error} (R_E)\) : (9)

\[
R_E = \frac{|y_i - \hat{y}_i|}{y_i} i = 1,2,3,4,5,6
\]

(9)

Calculate \(MRE\) (Mean Relative Error):

\[
MRE = \frac{1}{6} \sum_{i=1}^{6} R_E_i
\]

(10)

Solve for as shown in Table 2:

**Table 2.** Fitted equations and MRE data for different categories

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

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

\[
y_{ij} = C_{ij}(1 + r_{ij})
\]

(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, .

\[
C_{ij} = \frac{F_{ij} + V_{ij}}{Q_{ij}}
\]

(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.

\[
r_{ij} = \frac{S_{ij} - W_{ij}}{W_{ij}}
\]

(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:

\[
y_{ij} = \frac{(W_{ij}Q_{ij} + y_{ij}(1 - L_{ij}))}{x_{ij}} (1 + r_{ij})
\]

(14)

The profit maximization is expressed as follows [7]:

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

(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:

1. Type 1 expected sales volume, \( q_{ij} < 15 \text{ kg} \), markup rate \( 155 \leq r_{ij} \leq 165 \) (most of the superstores are low-market, high-profit goods with higher profit per unit of goods)

2. 2nd expected sales volume, \( q_{ij} < 45 \text{ kg} \), markup rate \( 135 \leq r_{ij} \leq 145 \) (medium-volume merchandise that guarantees sales while earning more profit)

3. Expected sales volume, \( q_{ij} \geq 45 \text{ kg} \), markup rate \( 115 \leq r_{ij} \leq 125 \) (high volume goods, thin margins through lower prices)

Based on the linear relationship, the following constraint is constructed:

\[
y_i = a_{i0} + a_{i1} x_i i = 1,2,3,4,5,6
\]

The planning model is as follows: [8]

\[
\max_{q_{ij}} (p_{ij} - w_{ij}) - [W_{ij}Q_{ij} + P_{ij}(1-L_{ij})] = y_{ij} (1 + r_{ij})x_{ij}
\]

subject to:

\[
y_{ij} = (1 + r_{ijk})c_{ij} \]

\[
x_{ij}, \quad w_{ij} \geq 0
\]

\[
155 \leq r_{ij1} \leq 165\%
\]

\[
135 \leq r_{ij2} \leq 145\%
\]

\[
115 \leq r_{ij3} \leq 125\%
\]

\[
j = 1,2\ldots,7
\]

\[
i = 1,2\ldots,6
\]

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

<table>
<thead>
<tr>
<th>dates</th>
<th>cauliflower (Brassica oleracea var. botrytis)</th>
<th>philodendron</th>
<th>chilli</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total daily replenishment</td>
<td>pricing strategy (%)</td>
<td>Total daily replenishment</td>
</tr>
<tr>
<td>2023-07-01</td>
<td>56</td>
<td>53</td>
<td>223</td>
</tr>
<tr>
<td>2023-07-02</td>
<td>20</td>
<td>24</td>
<td>251619</td>
</tr>
<tr>
<td>2023-07-03</td>
<td>43</td>
<td>48</td>
<td>272</td>
</tr>
<tr>
<td>2023-07-04</td>
<td>39</td>
<td>89</td>
<td>156</td>
</tr>
<tr>
<td>2023-07-05</td>
<td>51</td>
<td>43</td>
<td>272</td>
</tr>
<tr>
<td>2023-07-06</td>
<td>33</td>
<td>42</td>
<td>260</td>
</tr>
<tr>
<td>2023-07-07</td>
<td>28</td>
<td>55</td>
<td>199019</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dates</th>
<th>eggplant</th>
<th>edible mushroom</th>
<th>Aquatic rhizomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total daily replenishment</td>
<td>pricing strategy (%)</td>
<td>Total daily replenishment</td>
</tr>
<tr>
<td>2023-07-01</td>
<td>10</td>
<td>93</td>
<td>67</td>
</tr>
<tr>
<td>2023-07-02</td>
<td>9</td>
<td>39</td>
<td>70</td>
</tr>
<tr>
<td>2023-07-03</td>
<td>5544</td>
<td>62</td>
<td>30</td>
</tr>
<tr>
<td>2023-07-04</td>
<td>10</td>
<td>101</td>
<td>52</td>
</tr>
<tr>
<td>2023-07-05</td>
<td>3</td>
<td>100</td>
<td>44</td>
</tr>
<tr>
<td>2023-07-06</td>
<td>6965</td>
<td>104</td>
<td>60</td>
</tr>
<tr>
<td>2023-07-07</td>
<td>6</td>
<td>80</td>
<td>59</td>
</tr>
</tbody>
</table>
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

\[ \max \sum_i (\beta_i - \eta_i) x_i \]  

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].

\[ \text{Maximize } Z = \sum (\beta_i - \eta_i) z_j \]
\[ \text{s.t. } \begin{cases} 
27 \leq \sum z_i \leq 33 \\
z_i \geq 2.5 \forall i \\
p_i \leq 100\% \forall i
\end{cases} \]

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.

\[ \begin{cases} 
\max \sum_i (\beta_i - \eta_i) x_{i\text{average}} \\
p_{\text{average}} \\
\beta_{\text{average}} \\
\eta_{\text{average}}
\end{cases} \]

(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.

\[ Z = -\eta_B + \sum \eta_j z_j \]
\[ z_B = B - \sum A_i x_j \]

where \( z_B \) is the coefficient of the objective function of the basic variable, \( \eta_j \) is the coefficient of the objective function of the non-basic variable, \( p_i \) is the value of the basic variable, \( \beta_i \) is the value of the the values of the right-hand side of the constraints, and \( A_i \) is a column of the coefficient matrix.

(4) Checking the optimality condition: check if there are non-negative \( \eta_j \) in the rows of the objective function. If all \( \eta_j - \eta_B \leq 0 \), 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>
<thead>
<tr>
<th>product code</th>
<th>Total daily replenishment (KG)</th>
<th>Pricing strategy (%)</th>
<th>product code</th>
<th>Total daily replenishment (KG)</th>
<th>Pricing strategy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>102900011034330</td>
<td>10</td>
<td>53</td>
<td>102900011034439</td>
<td>5</td>
<td>59</td>
</tr>
<tr>
<td>102900005115779</td>
<td>10</td>
<td>134</td>
<td>102900011035740</td>
<td>1</td>
<td>86</td>
</tr>
<tr>
<td>102900005115878</td>
<td>9</td>
<td>142</td>
<td>102900011013274</td>
<td>2</td>
<td>125</td>
</tr>
<tr>
<td>102900011021842</td>
<td>7</td>
<td>71</td>
<td>102900011023464</td>
<td>5</td>
<td>128</td>
</tr>
<tr>
<td>102900005115861</td>
<td>18</td>
<td>94</td>
<td>102900011030059</td>
<td>13</td>
<td>98</td>
</tr>
<tr>
<td>102900005116530</td>
<td>18</td>
<td>47</td>
<td>102900011030097</td>
<td>9</td>
<td>94</td>
</tr>
<tr>
<td>102900011030905</td>
<td>1</td>
<td>63</td>
<td>102900011030110</td>
<td>15</td>
<td>138</td>
</tr>
<tr>
<td>102900005116899</td>
<td>25</td>
<td>128</td>
<td>102900011008164</td>
<td>6</td>
<td>36</td>
</tr>
</tbody>
</table>

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


