Vegetable Pricing and Restocking with Markov Models

Yanhe Zhu *, Qiaoman Xie, Gege Zhang

Shaanxi Normal University, Xian, China, 710119

* Corresponding Author Email: Zyh14370315@outlook.com

Abstract. The freshness period of most vegetables in fresh produce superstores is relatively short, thus solving the vegetable replenishment and pricing decision problem is of practical significance. In order to make the best decision on vegetable commodity replenishment and pricing, this paper firstly fits a linear fit to the total sales volume and cost-plus pricing. Secondly, an optimisation model is established with maximum revenue as the objective function, cost-plus pricing and inbound unit price as the decision variables, and replenishment volume equal to sales volume as the constraint after considering the wastage rate. Finally, cost-plus pricing and in-stock unit price are predicted using cubic exponential smoothing.The utilization of Markov models allows for short-term forecasting based on a limited amount of recent data and possesses generality.

Keywords: Optimized Model, Cubic Exponential Smoothing, Markov model.

1. Introduction

In a fresh supermarket, vegetable products have the characteristic of a short shelf life, deteriorating in quality as time passes, and most varieties cannot be sold the next day if unsold. Therefore, the supermarket replenishes its stock daily based on historical sales and demand data. Given the diversity of vegetable categories, their different sources of origin, and the fact that purchases are typically made in the early morning, retailers must formulate replenishment plans for vegetable products without knowing the specific items and purchase prices. Cost-plus pricing is often used for such products, and damaged goods are usually sold at a discount. Reliable market demand analysis is crucial for replenishment and pricing decisions. On the demand side, vegetable sales are time-dependent; on the supply side, the large variety of items available from April to October requires a rationalisation of the sales mix within the limited selling space of the supermarket.

To make optimal decisions regarding daily replenishment quantities and pricing for vegetables, current research efforts by domestic and international scholars have predominantly focused on price forecasting, yielding significant achievements [1-3]. For instance, scholars such as Wang Yujie [4]have successfully employed the ARIMA model to forecast garlic prices based on Chinese garlic price data spanning from January 2004 to September 2021. Peng Hongxing [5]and other scholars have separately utilized BP (Backpropagation), LSTM (Long Short-Term Memory), and ARIMA models to forecast vegetable prices, with findings indicating that the ARIMA model exhibits superior performance in predicting vegetable prices compared to the other two models. In addition, scholars like Liu Hebing [6] have constructed a comprehensive agricultural product price composite forecasting model based on wavelet transformation and BP neural networks. This model has achieved a high level of accuracy in monthly price forecasts for five different types of vegetables.However, the aforementioned models all require a substantial amount of data to make accurate predictions and are not suitable for situations where there is insufficient known data. In contrast, the Markov model employed in this study enables predictions with a smaller amount of data.

Based on the sales data provided by the Higher Education Press for a specific supermarket, this paper examines the relationship between the total sales volume of vegetable categories and cost-plus pricing and provides daily replenishment quantities and pricing strategies for the upcoming week (July 1st to 7th, 2023). Firstly, this paper takes the time scale of 1st June to 31st July for each year from 2020 to 2023 and investigates the total sales volume of each vegetable category as the dependent variable and cost-plus pricing as the independent variable. The data is fitted using the Polyfit function in Matlab, resulting in an equation describing their relationship. Secondly, with the objective of maximizing revenue, cost-plus pricing and purchase unit price are treated as decision variables, while the replenishment quantity equals the sales quantity is imposed as a constraint. The paper utilizes the cubic exponential smoothing method [7]to forecast the cost-plus pricing and purchase unit price for July 1st to 7th, 2023, thus determining the sales volume and replenishment quantity.

Due to the complex and dynamic nature of real-world conditions and the multitude of influencing factors, this paper, while operating under certain constraints, formulates a single-item replenishment plan as per the supermarket's intentions. It stipulates that the total number of saleable items should fall within the range of 27 to 33, while simultaneously requiring that the replenishment quantities for each item meet a minimum display requirement of 2.5 kilograms. Based on the available assortment from 24th to 30th June 2023, the replenishment quantities and pricing strategy for individual items for 1st July 2023 are provided. This can be viewed as a planning problem with an objective function of maximum revenue. The constraints are that the total number of items available for sale ranges from 27 to 33, the quantity ordered for each item satisfies a display quantity of not less than 2.5, and the quantity of each item stocked after taking into account the rate of wastage is equal to the quantity sold. The decision variables include the purchase quantities, purchase unit prices, and cost-plus pricing for individual items. Utilizing the Markov model [8-9], the study predicts the numerical values of the decision variables for July 1st and subsequently determines the maximum revenue achievable.

2. Main content

2.1. Vegetable replenishment and pricing decision based on cubic exponential smoothing method

Due to the seasonal nature of vegetable product sales, for better forecasting the daily replenishment quantities and cost-plus pricing for July 1st to 7th, 2023, where the value of cost-plus pricing is taken as the average pricing for each vegetable category, this paper selects the time frame from June 1st to July 31st for each year from 2020 to 2023. The establishment of the model in this paper will unfold in two parts: firstly, the analysis of the relationship between the total sales volume and cost-plus pricing for various vegetable categories; and secondly, the formulation of replenishment and pricing decisions.

2.1.1 Analysis of the relationship between total sales and cost-plus pricing for each vegetable category

The study of this part is the study of the relationship between univariate and univariate; accordingly, this paper chooses the Polyfit function in Matlab to fit the two linearly. Taking cost-plus pricing as the independent variable and total sales volume as the dependent variable, the relationship between the sales volume Q_i for each category and its cost-plus pricing P_i is obtained as follows:

$$Q_1 = -12.66P_1 + 262.1 \tag{1}$$

$$Q_2 = -2.633P_2 + 63.5 \tag{2}$$

$$Q_3 = -2.534P_3 + 63.71 \tag{3}$$

$$Q_4 = -0.9634P_4 + 29.95 \tag{4}$$

$$Q_5 = -1.371P_5 + 98.52 \tag{5}$$

$$Q_6 = -6.442P_6 + 148.4 \tag{6}$$

The fitted graphs for example Q_1 and Q_6 are depicted as follows in Figures 1 and 2, respectively.

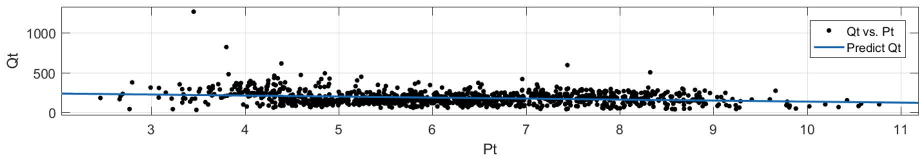

Figure 1. The fitted graphs for Q_1 and P_1

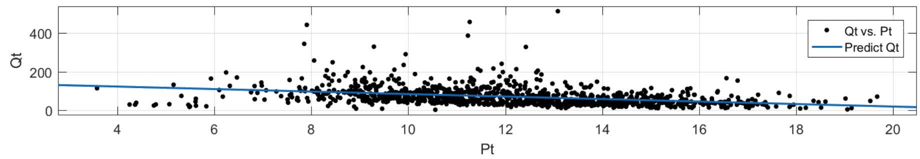

Figure 2. The fitted graphs for Q_6 and P_6

In connection with the law of demand can be seen[10], in general the higher the price of the commodity the lower the demand, and the above six relationships are consistent with this law, indicating that the results of this prediction has a certain reference value.

2.1.2 Modelling and solving replenishment and pricing models

1) Determining the three elements

Incoming unit price $G_i(t)$ (averaged over all individual items in each category) and cost-plus pricing $P_i(t)$ are chosen as decision variables, and the objective function is $\max W$. If the replenishment volume is less than the sales volume, there is a group of consumers who are not able to realise the purchase and therefore lose part of the potential revenue. Conversely, if the replenishment quantity exceeds the sales quantity, some vegetables may not be sold in a timely manner, and their costs cannot be realized. Therefore, this paper argues that the total revenue W can be maximised if and only if the replenishment volume F_i is the same as the sales volume Q_i after considering the wastage rate B_i. Thus, the constraint condition is that the replenishment quantity after considering the loss rate equals the sales quantity. To reduce the amount of data processing, this paper takes the average of the loss rates for each product category as the loss rate for that category.

2) Model formulation

$$\max W = \sum_{t=1}^{7}\sum_{t=1}^{6}\left[Q_i(t)\cdot P_i(t) - F_i(t)\cdot G_i(t)\right] \tag{7}$$

$$st.\frac{F_i(t)}{1+B_i} = Q_i(t) \tag{8}$$

In which, $Q_i(t)$ satisfies the relationship between Q_i and P_i described by equations (1.1) to (1.6) in Section 2.1.1

3) Model solving

Based on the aforementioned model, it is evident that the core of this section is the prediction of $G_i(t)$ and $P_i(t)$ for July 1st to 7th, 2023. Since data closer to the present time is more valuable, this paper employs the triple exponential smoothing method to forecast $G_i(t)$ and $P_i(t)$.

a) Modelling of the cubic exponential smoothing method

$$\hat{y}_t + T = a_t + b_t T + c_t T^2, T = 1, 2... \tag{9}$$

$$\text{In which}\ \begin{cases} a_t = 3S_t^{(1)} - 3S_t^{(2)} + S^3{}_t \\ b_t = \dfrac{a}{2(1-a)^2}\left[(6-5a)S_t^{(1)} - 2(5-4a)S_t^{(2)} + (4-3a)S_t^{(3)}\right] \\ c_t = \dfrac{a^2}{2(1-a)^2}\left[S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}\right] \end{cases} \tag{10}$$

$$\begin{cases} S_t^{(1)} = \alpha y_t + (1-\alpha) S_{t-1}^{(1)} \\ S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S^2{}_{t-1} \\ S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha) S^2{}_{t-1} \end{cases} \tag{11}$$

b) Cubic exponential smoothing method of model solving.

Matlab was used to predict the sales volume, cost-plus pricing, purchase volume and purchase unit price of the six vegetable categories, and the difference between the predicted and original values was obtained by comparing them on the image. According to formula $R^2 \equiv \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \Big/ \sum_{i=1}^{n}(y_i - \bar{y})^2$, the goodness of fit for each category is shown in Figure 3. The goodness of fit for the six vegetable categories is shown in Table 1. Therefore, it can be considered that the forecasts for July 1st to 7th, 2023, are relatively accurate. By incorporating the predicted values into the model, the maximum total revenue is calculated to be 12,315.68 yuan.

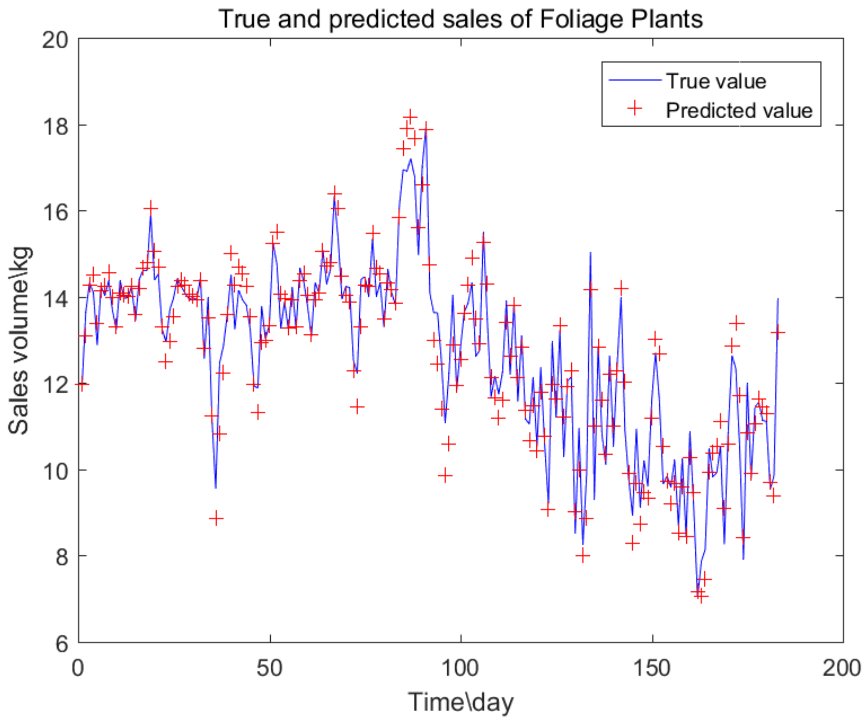

Figure 3. Comparison of real and forecast sales of foliage plants category.

Table.1. Goodness-of-fit for six major vegetable categories.

Vegetable Categories	Goodness of Fit
Foliage Plants	0.876795068
Flowering Cauliflower Category	0.906127699
Aquatic Rhizomatous Plants	0.932024109
Solanaceous Vegetables	0.888221237
Chili peppers	0.908638636
Edible Mushrooms	0.894943858

2.2. Vegetable Replenishment and Pricing Decisions Based on the Markov Model

1) Determining the three elements

In this section, the decision variables include the purchase quantities $f_i(t)$ for individual items, cost-plus pricing $h_i(t)$, and purchase unit prices $g_i(t)$.The objective function is to maximise revenue $maxW$. Based on this section and 2.1.2, the constraints are that the total number of items available for sale ranges from 27 to 33, the order quantity of each item satisfies the display quantity of not less than 2.5, and the purchase quantity of each item is equal to the sales quantity after considering the wastage rate.

2) Model formulation

$$maxW = \sum_{i=1}^{251}\left[q_i(t) \cdot h_i(t) - f_i(t) \cdot g_i(t)\right] \tag{12}$$

$$st.\begin{cases} 27 \le \sum_{i=1}^{251} x_i \le 33 \\ \dfrac{f_i(t)}{1+b_i(t)} \ge 2.5 \quad \text{or} \; \dfrac{f_i(t)}{1+bi(t)} = 0 \\ \dfrac{f_i(t)}{1+bi(t)} = qi(t) \end{cases} \tag{13}$$

In which, $X_i = \begin{cases} 1 & \text{The ith individual item is available for sale.} \\ 0 & \text{Unsellable without purchasing the ith item} \end{cases}$

3) Model solving

The key to the unit sales price solution here lies in the accurate prediction of the 1 July 2023 $f_i(t)$, $h_i(t)$ and $g_i(t)$. Since this paper is based on the data from 24-30 June 2023, the prediction of the data on 1 July, that is, the use of recent data on the prediction of future conditions. Secondly the Markov model has the function of predicting the future based on a small amount of existing data, so this paper chooses to build the model to make predictions for $f_i(t)$, $h_i(t)$, $g_i(t)$.

a) Markov modelling

$S(n)$ denotes the state vector of the predicted object at moment n and p is the one-step transfer probability matrix. Further prediction of the state vector $S(n+m)$ at moment $n+m$ has: $S(n+m) = S(n) \cdot P^m$

In which,

$$P = (p_{ij}) = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{14}$$

Denote the conditional probability $p^{(m)}{}_{ij} = P\{X_{n+m} = j \mid X_n = i\}$ as the m-step transfer probability of the Markov chain, and correspondingly, $P^{(m)} = \left[p_{ij}^{(m)}\right]$ is referred to as the m-step transition matrix.

b) Solution of the Markov model

The average transfer rate matrix can be obtained using Matlab as:

$$P=\begin{pmatrix} 0.859 & 0 & \cdots & 0.000 & 0.000 \\ 0 & 0.358 & \cdots & 0.000 & 0.000 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0.000 & 0.000 & \cdots & 0.837 & 0.000 \\ 0.000 & 0.000 & \cdots & -1.038 & 13.114 \end{pmatrix}_{49\times 49} \tag{15}$$

Consequently, it becomes possible to forecast the purchase quantities $f_i(t)$, selling prices $h_i(t)$, and purchase unit prices $g_i(t)$ for each individual item on July 1st. The maximum revenue is determined to be 632.074 yuan. Selected predicted values for decision variables are presented in Table 2.

Table.2. Prediction for selected decision variables

Selection of rankings	Corresponding Product Number	Individual product name	Sales Volume (Kilograms)	sale Price	Selection of rankings	Corresponding Product Number
3	102900005115779	Yunnan lettuce	33.507	9.200	38.616	5.738
4	102900005115786	Bamboo Leaf Vegetable	14.243	3.800	16.183	2.319
5	102900005115823	Shanghai Choy	3.111	8.000	3.559	4.121
…	…	…	…	…	…	…
46	102900051000463	Round aubergine (2)	1.406	7.067	1.500	4.096

3. Conclusions

The freshness period of most vegetables in fresh produce superstores is relatively short, thus solving the problem of vegetable replenishment and pricing decision making. This study is based on available data and focuses on researching replenishment and pricing decisions for vegetable products. Firstly, the data is preprocessed by normalization. Secondly, using the "polyfit" function in MATLAB, linear fitting is applied to the sales total and cost-plus pricing. Subsequently, with the objective of maximizing revenue, cost-plus pricing and purchase unit price are treated as decision variables. The model considers replenishment quantities, accounting for losses, to be equal to sales quantities as a constraint. Finally, the study utilizes triple exponential smoothing to forecast the cost-plus pricing and purchase unit price for July 1st to 7th, 2023. The calculation yields a maximum total revenue of 12,315.68 yuan over seven days.

In the presence of constraint conditions, the study initially treats purchase quantities, cost-plus pricing, and purchase unit price for individual items as decision variables, with the objective of maximizing revenue. Secondly, the optimisation model is established with the constraints of limiting the total number of items available for sale, ordering quantity display requirement of each item, and the purchase quantity of each item equals to the sales quantity after considering the wastage rate. Finally, the Markov model is employed to predict purchase quantities, cost-plus pricing, and purchase unit price for individual items on July 1st, 2023, resulting in a maximum revenue of 632.074 yuan.

References

[1] Cheng Jinpeng. Research on vegetable price forecasting in Henan Province based on ARIMA-BP neural network [D]. Henan Agricultural University,2019.

[2] YU Yongsong,PANG Zhengwu,ZHOU Yenning et al. Establishment of grey prediction model GM(1,1) for vegetable yield in Guangxi and its correlation analysis [J]. Guangdong Agricultural Science,2018,45(07):157-164.

[3] AHINLI M A. Potato price forecasting with Holt-Win-ters and ARIMA methods: A case study [J]. American Merican Journal of Potato Research, 2020, 97(4):336-346.

[4] WANG Yujie, BAI Li, ZHAO Banghong. Garlic price forecast in China based on ARIMA model [J]. Vegetable, 2021(12):50-54.

[5] PENG Hongxing, ZHENG Kaihang,HUANG Guobin et al. Vegetable price prediction based on BP, LSTM and ARIMA models [J]. Chinese Journal of Agricultural Mechanical and Chemical Sciences, vol. 2020,41(04):193-199.

[6] LIU Hebing, HAN Jingjing, XI Lei. Research on price prediction of agricultural products by wavelet transform-BP neural network [J]. China Agricultural Information. 2019,31(06):85-92.

[7] J. Zhang, J. Zhu. Prediction of total waste volume based on multiple regression and cubic exponential smoothing [J]. Journal of Qilu University of Technology. 2020,34(04):69-74.

[8] Yan Maode,Jun Mengmeng,Zuo Lei. Optimal energy consumption control of vehicle queues under Markov chain communication topology [J/OL]. Control and Decision,1-9 [2023-10-25].

[9] Yan YF. Matlab-based Markov chain prediction optimisation model and its application [J]. Integrated Circuit Applications,2022,39(07):268-270.

[10] Gao Hongye. Western economics (7th edition) [M]. Beijing:People's University of China Press,2017.