Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Study on Pricing and Replenishment Decision of Vegetable Products Based on SARIMA Model

Feifan Yu [1, #], Lei You [1, #], Xinyue Han [2], Haoze Ma [2], Shijun Sun [1], Liyin Zhang [2], Yuesheng Zhao [3, *]

[1] Tus College of Digit, Guangxi University of Science and Technology, Liuzhou, China, 545006

[2] School of International Education of Guangxi University of Science and Technology, Guangxi University of Science and Technology, Liuzhou, China, 545006

[3] College of Automation, Guangxi University of Science and Technology, Liuzhou, China, 545006

* Corresponding author: zhyuesheng@126.com

[#]These authors contributed equally.

Abstract. In fresh food superstores, vegetable commodities are characterized by a short shelf life and deterioration in quality over time. To meet market demand and profit, merchants usually sell vegetables with shipping losses and deteriorating quality at a discount. Therefore, the pricing and replenishment strategies of vegetables have an important impact on the profit maximization of fresh produce superstores. The aim of this paper is to reveal the patterns by processing and analyzing the past sales data. The distribution pattern between different categories of vegetables is obtained through Pearson's coefficient. Then the correlation coefficients between different vegetable categories were calculated by ridge regression model. In the order of aquatic root vegetables, cauliflower vegetables, leafy vegetables, pepper vegetables, eggplant vegetables and edible mushroom vegetables, the coefficients of sales unit price were -0.535, -0.936, -2.498, -0.658, -0.279 and -1.157 respectively. Then the optimal model and the daily replenishment and pricing strategy for maximizing the revenue were determined by using the SARIMA model. The above model helps the superstore to adjust the replenishment plan and pricing strategy, and to better meet the market demand and increase sales revenue and profit. Also, it can be extended from a specific problem to other products with similar characteristics.

Keywords: Heat map; Ridge Regression; SARIMA.

1. Introduction

In the fresh produce retail sector, vegetables are characterized by short shelf lives and deteriorating quality over time. To meet market demands and ensure profitability, retailers often resort to discounting vegetables that suffer from transportation losses and quality deterioration. Although the spring and summer seasons offer a wide variety of vegetable options, subjective consumer behavior complicates quality assurance in the fresh produce market, potentially leading to unsold inventory risks [1]. In the operation of fresh produce markets like vegetables, asymmetric information during procurement transactions makes it challenging for retailers to have a clear understanding of specific item prices and procurement costs. Thus, market demand analysis is needed to formulate corresponding replenishment and pricing decisions [2].

There have been some research achievements in related fields, with some studies focusing on online retailers' efforts to stimulate consumer demand through green value-added initiatives. These studies have established joint decision-making models for green distribution technology investment, dynamic pricing, and replenishment, aiming to maximize total profits per unit time. By applying the Pontryagin maximum principle, researchers have identified optimal strategies for green distribution technology investment, dynamic pricing, and replenishment, and conducted numerical examples and sensitivity analyses of key system parameters. The results indicate that under different perishability rates and green distribution technology decay rates, the investment and replenishment quantities of green distribution technology in enterprises vary, but pricing strategies remain unaffected [3]. On the other hand, other studies have focused on the dynamic pricing and inventory planning issues in the fresh food supply chain. These studies have considered factors such as time, price, and freshness in affecting the demand for traditional channels and community-based e-commerce channels for fresh food and have established dynamic equilibrium models for suppliers and retailers based on differential variational inequalities. The research reveals that within the sales cycle of fresh food products, there exists a dynamic pricing breakpoint, allowing suppliers and retailers to adopt different pricing strategies based on the changing quality characteristics of products to enhance profitability. Additionally, they have proposed dynamic pricing discount strategies for the dual-channel fresh food supply chain to reduce losses from expired products [4].

Previous studies have focused more on theoretical derivation or model construction, lacking support from actual sales data. This may result in significant deviations between their research results and the actual situation. Moreover, they did not consider the influence of actual sales data in the modeling process or failed to fully utilize available data information. Lastly, their studies lack specific pricing and replenishment strategy recommendations, thus failing to provide practical guidance for retailers. In contrast, this paper reveals patterns in vegetable sales through the analysis of historical sales data, making the research more aligned with reality and enhancing its credibility and applicability. By employing quantitative analysis methods such as Pearson coefficients and ridge regression models, it uncovers pricing influencing factors among different vegetable categories, emphasizing an intuitive understanding of sales. Furthermore, it utilizes the SARIMA model to determine optimal replenishment quantities and pricing strategies, providing practical replenishment plans and pricing strategies for retailers [5].

2. The Principle of Model Establishment

2.1. The principle of Ridge Regression Model

The Ridge Regression model is a linear regression technique used to address multicollinearity issues, particularly applicable in analyzing sales data of vegetables in fresh food supermarkets. Its main objective is to enhance the stability and accuracy of the model to better understand the patterns in vegetable sales. The model's principle is based on improving Ordinary Least Squares (OLS) by adding a regularization term, known as the ridge penalty term, to the loss function to tackle multicollinearity problems. Through the Ridge Regression model, you can address multicollinearity issues present in vegetable sales data, making the model more stable and reliable. The introduction of the ridge penalty term reduces the coefficient estimates of the model, thereby reducing variance, preventing overfitting, and partially suppressing the influence of unimportant variables on the model. This allows for a more accurate analysis of the pricing factors among different categories of vegetables, providing supermarkets with more reasonable replenishment plans and pricing strategies.

Specifically, the core principle of the Ridge Regression model is to adjust the coefficient of the penalty term (ridge parameter) to control the complexity of the model, thereby balancing the model's goodness of fit and generalization ability. The loss function of the Ridge Regression model can be represented as:

$$L(\beta) = \sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}\beta_j^2 \tag{1}$$

Where y_i is the sales price of vegetable products in the fresh food supermarket, x_{ij} is the $j-th$ feature of the $i-th$ vegetable product, β_j is the coefficient corresponding to feature j of the vegetable product, indicating the degree of influence of this feature on the sales price, and λ is the penalty coefficient in the Ridge Regression model, used to control the complexity of the model and prevent overfitting.

2.2. The principle of SARIMA Model

The SARIMA (Seasonal Autoregressive Integrated Moving Average) model is a statistical model used for time series analysis and forecasting. Its purpose is to capture features such as seasonality, autoregression, and moving averages in time series data to make predictions about future trends and seasonal patterns.

The model is based on the principles of the ARIMA (Autoregressive Integrated Moving Average) model but incorporates seasonal factors present in time series data. Its mathematical expression is as follows [6]:

$$(1-\phi_1 L-\ldots-\phi_p L^p)(1-\Phi_1 L^s-\ldots-\Phi_p L^{ps})(1-L)^d(1-L^s)^D y_t=(1+\Theta_1 L+\ldots+\Theta_Q L^{Qs})\varepsilon_t \tag{2}$$

Where y is the sales volume or sales sales number of vegetables. p,d,q are the respective orders of seasonal autoregression, differencing, and moving average. P,D,Q are the respective orders of seasonal autoregression, seasonal differencing, and seasonal moving average. S denotes the length of the seasonal cycle, and ε_t is the error term of the time series data.

The workflow of the SARIMA model includes the following three steps:

(1) Model identification and parameter determination: By analyzing autocorrelation and partial autocorrelation plots, the orders of autoregression (AR) and moving average (MA), as well as the length of the seasonal cycle in the SARIMA model, are determined. This step is crucial to ensure that the selected model effectively captures the characteristics of the time series data.

(2) Parameter estimation and model fitting: Utilizing methods such as maximum likelihood estimation, the selected SARIMA model is parameterized. Subsequently, the estimated parameter values are inserted into the model for fitting, aiming to capture the seasonal, autocorrelation, and moving average features in the time series data.

(3) Model diagnosis and prediction: The fitted SARIMA model undergoes residual analysis and model diagnostics to ensure its effectiveness and accuracy. Finally, the well-fitted model is employed to forecast future time series data, evaluating the accuracy and reliability of the forecasted results. The workflow diagram is shown in Figure 1.

Figure 1. Flow chart of SARIMA model

3. Results

3.1. Data Preprocessing

The vegetables mentioned in this paper are of various categories, so the "category name" is quantified as 0, 1, 2, 3, 4, 5 for foliage, cauliflower, aquatic rhizome, eggplant, pepper, and edible fungus, respectively. The data is sourced from the problem statement of the 2023 National College Student Mathematical Modeling Competition, hosted by Gaokao Society.

In order to better analyze the relationship between the characteristics, this paper creates new variables for the number of orders for a single item as well as the total sales. Based on the date of sale, the number of daily single product orders as well as the total daily sales (total sales = sales × unit price) are calculated to obtain the number of orders, total sales and average sales for the following classification names. The processed data is shown in the following table 1:

Table 1. Daily sales data by category

Serial number	Date of sale	Classification name	Total daily sales	Average sales
0	2020-7-1	Aquatic rhizomes	70.2838	8.857681
1	2020-7-1	philodendron	1503.7896	4.113706
2	2020-7-1	Cauliflower	592.53	5.327872
3	2020-7-1	eggplant	176.818	2.920933
4	2020-7-1	capsicum	759.9902	3.315201
5	2020-7-1	edible mushroom	368.602	3.188564
6	2020-7-1	Aquatic rhizomes	53.1208	6.5204
…	…	…	…	…

3.2. Analysis of the interrelationships between different categories of vegetables

In order to gain insight into the degree of correlation between different vegetable categories, this paper redefines and integrates the vegetable sales data of different categories by constructing a pivot table, which results in a correlation matrix between each vegetable category. Pearson's correlation coefficient is further utilized to generate a heat map showing the correlation between vegetable categories [7]. From Figure 2, we can get that there is a strong positive correlation between edible mushroom vegetables and chili vegetables, and their trends are similar. A relatively strong positive correlation exists between cauliflower vegetables and leafy vegetables, pepper vegetables and aquatic root vegetables, pepper vegetables and cauliflower vegetables, edible mushrooms vegetables and aquatic root vegetables, edible mushrooms vegetables and cauliflower vegetables. Whereas, very weak positive correlations existed between eggplant vegetables and aquatic root vegetables, edible mushroom vegetables and eggplant vegetables.

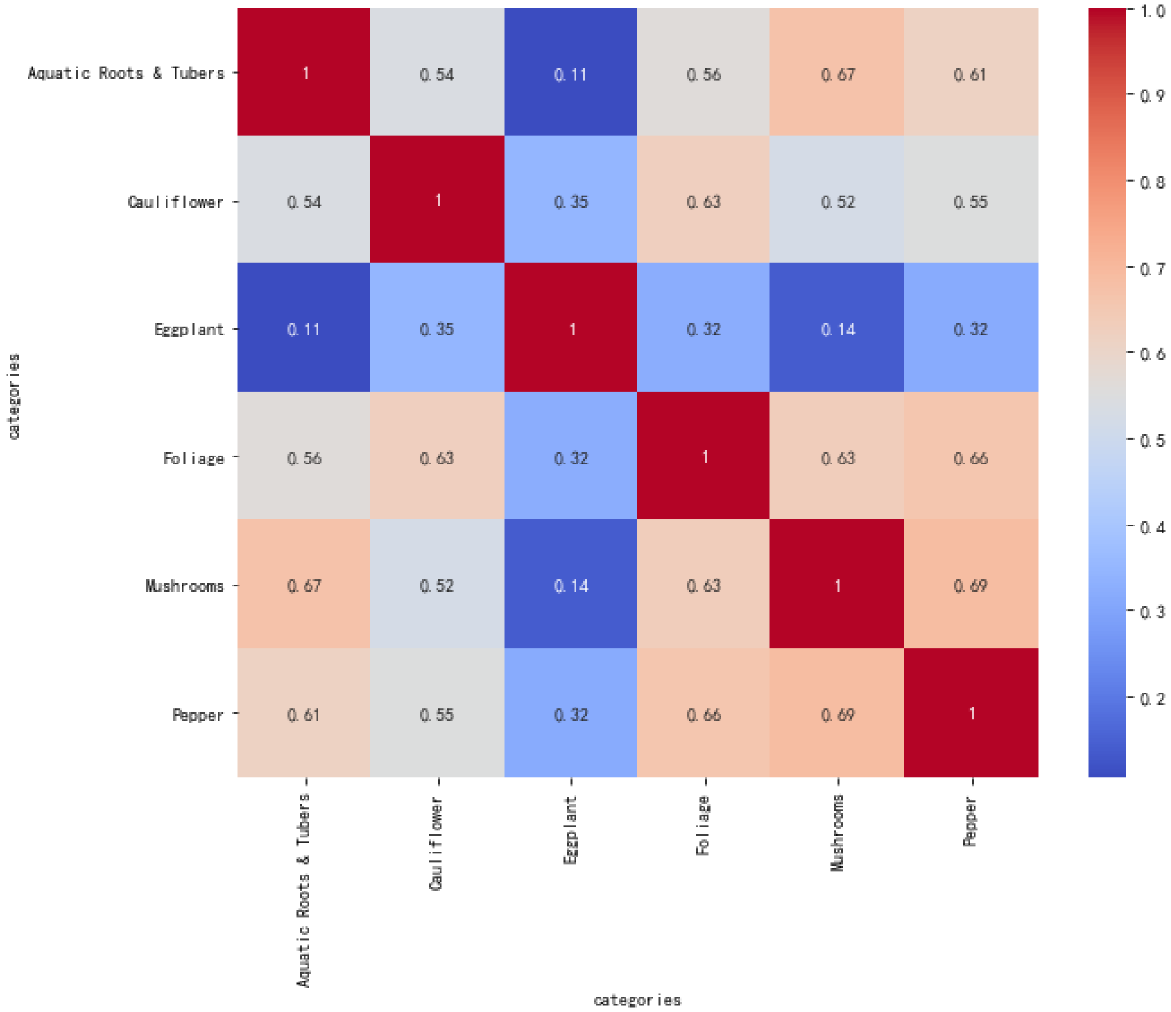

Figure 2. Vegetable Category Interrelationship Analysis

3.3. Calculation of unit price coefficient

Based on the actual sales volume of each category and the cost-plus pricing of each category, combined with the ridge regression model can be obtained, as shown in Table 2:

Table 2. Results of ridge regression analysis

K=0.16	Non-std Coeff	R^2	Adjustment R^2
	B		
Constant	4.896	0.972	0.972
Number of Orders	0.069		
Total Sales	0.033		
Average Sales	0.398		
Sales (kg)	0.731		
Unit Price (RMB/kg)	-0.535		

Taking aquatic root vegetables as an example, the results of the above ridge regression on the indicators of aquatic root vegetables show that the model's goodness-of-fit R^2 is 0.972, and the model performance is excellent. The coefficient of unit sales price is -0.535. According to this method, the ridge regression analysis of other categories of vegetables shows that the goodness of fit R^2 of cauliflower vegetables, foliage vegetables, pepper vegetables, eggplant vegetables, edible mushroom vegetables are 0.970, 0.972, 0.973, 0.895, 0.974. The coefficients of unit sales price are -0.936, -2.498, -0.658, -0.498, -0.498, -0.498, -0.936, -0.658, and -0.498 respectively, -0.658, -0.279, -1.157.

3.4. Evaluation of the SARIMA Model

Based on the inspection of the time series data's plot, autocorrelation function (ACF), and partial autocorrelation function (PACF), the values of p, d, q, P, D, Q, and s for the model are estimated. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are computed [8]. The results are presented in Table 3 below:

Table 3. Model evaluation

SARIMA model (5, 1, 5)×(0, 0, 1, 12)		
term	notation	value
sample size	N	1085
Q-statistic	Q6 (p-value)	0.489
	Q12 (p-value)	0.73
	Q18 (p-value)	0.72
	Q24 (p-value)	0.577
	Q30 (p-value)	0.587
Information guidelines	AIC	11451.775
	BIC	11511.636
goodness of fit	R2	0.608

From the model evaluation table, the information criterion AIC is 11451.775, the smaller value of AIC shows that the model fits the data better to some extent. BIC is 11511. 636. The value of BIC is also smaller which means that the SARIMA model is still relatively good in balancing the effectiveness of the fit and the complexity of the model [9].

3.5. Forecasting Results of Daily Replenishment Quantity and Pricing Strategy

After completing the model examination, a white noise test is conducted. Based on the p-value of the Q statistic in the model evaluation table (p-value greater than 0.1 indicates white noise), the white noise of the model is tested. As shown in the table, all p-values are greater than 0.1, indicating that all statistics are white noise [10]. Finally, the predicted results of daily replenishment and pricing strategy are obtained, as shown in the table 4 and table 5.

Table 4. Forecast of daily replenishment of flowering and leafy vegetables

Number of steps (time)	Projected results
1	207.92658773653200
2	178.12890707832300
3	129.07320921260700
4	142.47099744383500
5	159.59895191313600
6	144.62980540729200
7	171.12939685003700

The order (time) is from day 1 to day 7, and the prediction result is the total daily replenishment of leafy and flowering vegetables. From the data in the table, it can be observed that the predicted replenishment varies between different days without a clear trend. The forecasted values for the first and seventh days are relatively high, while the prediction for the third day is lower.

Table 5. Forecast of pricing strategies for tomato vegetables

Number of steps (time)	Projected results
1	7.7431419821638100
2	7.7477257917501800
3	7.7498679125139200
4	7.7508689753373200
5	7.7513367952629600
6	7.7515554183880000
7	7.7516575860559200

The order (time) is from day 1 to day 7, and the prediction result is the pricing of eggplant vegetables. In comparison to the replenishment predictions, the forecasted pricing strategies show smaller variations and maintain a relatively stable trend. The pricing strategy for each day fluctuates between 7.74 and 7.75.

4. Conclusions and Outlook

This study explores the pricing and replenishment strategies of vegetable products in fresh food supermarkets using actual sales data and quantitative analysis methods. By analyzing historical sales data, we found that the pricing of different vegetable categories is influenced by various factors. We employed the Ridge Regression model and the SARIMA model to determine the optimal pricing and replenishment strategies. The results of the study indicate that implementing the proposed models can assist retailers in adjusting their replenishment plans and pricing strategies to better meet market demands, thereby increasing sales revenue and profit. This research provides practical guidance for the management of vegetable products and has both theoretical and practical significance.

Using the ridge regression model and SARIMA model to address pricing and replenishment strategies for vegetable products has the following advantages and disadvantages:

(1) The advantage of the ridge regression model lies in handling multicollinearity: Pricing and replenishment strategies for vegetable products are often influenced by multiple factors. The ridge regression model can effectively address multicollinearity among these factors, ensuring model robustness.

(2) The advantage of the SARIMA model lies in considering seasonality: Sales volume and revenue of vegetable products are typically influenced by seasonality. The SARIMA model can effectively capture this seasonality, providing more accurate predictions for pricing and replenishment decisions.

(3) To address the challenge of adjusting parameters in the ridge regression model, methods such as grid search or random search can be employed to find the optimal penalty coefficient, reducing the burden of manual parameter tuning. Given the difficulty in selecting parameters for the SARIMA model, the Auto-ARIMA algorithm can be used to automatically select the optimal ARIMA model parameters, reducing the complexity and subjectivity of parameter selection and improving the model's reliability.

References

[1] Yang Tianshan, Yuan Gonglin. Research on dynamic pricing and replenishment strategy of perishables considering investment in green distribution technology [J]. China Management Science, 2023, 31 (11): 279 - 287.

[2] Li Panfeng, Ma Zujun, Sun Hao. Research on blood supply and demand prediction based on SARIMA combined forecasting model[J]. Industrial Engineering and Management, 2023, 28 (03): 176 - 186.

[3] Zhang Jing, Liu Jifang, Zhou Xiangyang, et al. Analysis of the vegetable market situation in 2022 and outlook for 2023[J]. Chinese Vegetables, 2023, (01): 1 - 6.

[4] Yuan Jingting, Li Cuixia. Study on Dual-Channel Dynamic Pricing Mechanism of Dairy Products Considering Quality Loss and Random Demand [J]. Modern Management, 2023, 43 (06): 108 - 119.

[5] Yang N. The unique role in the solution of multicollinearity problems by ridge regression analysis [J]. Statistics and Decision Making, 2004 (3): 14 - 15.

[6] He Jinting, Chen Xingyan, Tao Tao, et al. Research on Furniture Order Demand Forecasting Method Based on SARIMA-BP Combination Model [J]. Furniture and Interior Decoration, 2024, 31 (02): 26 - 30.

[7] Zhou Xiangyu, Li Si. Short-term Forecasting of Express Delivery Volume in Jiangsu Province Based on TOPSIS Criteria and SARIMA Model [J]. Science and Technology & Industry, 2023, 23 (17): 136 - 142.

[8] Tang Jingyun, Xi Xin, Zhao Peng. Optimization of LSTM Model Based on Sales Forecasting [J]. Journal of Taiyuan Normal University (Natural Science Edition), 2024, 23 (01): 45 - 52.

[9] Liu Feng, Wang Rujing, Li Chuanxi. Application of ARIMA model in agricultural product price prediction [J]. Computer Engineering and Applications, 2009, 45 (25): 238 - 239+248.

[10] Fan Yingjie, Zhang Qing. Short-term prediction of heating energy consumption in green buildings based on time series autoregressive model [J]. Computer Measurement and Control, 2023, 31 (04): 289 - 294.