Study on the Pricing and Replenishment Strategies for Vegetable Commodities

Qinfa Zhuang[1], Kangbo Pei[2], Kaihua Wen[3], Fang He[4*]

[1]Business School, Shenzhen Technology University, Shenzhen 518118, China
[2]College of Health Science and Environmental Engineering, Shenzhen Technology University, Shenzhen 518118, China
[3]Future Technology School, Shenzhen Technology University, Shenzhen 518118, China
[4]College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518118, China

*Corresponding author's e-mail: hefang@sztu.edu.cn

Abstract. With the development of the economy and increased health awareness, the demand for vegetables continues to rise. Addressing the issue of short shelf life and price fluctuations in fresh supermarkets, this paper proposes a comprehensive optimization model aimed at optimizing inventory management, reducing food waste, and maximizing supermarket revenue through scientific pricing and replenishment strategies. Initially, correlation and time series analyses are conducted to establish a multiple linear regression model and a Seasonal Autoregressive Integrated Moving Average forecasting model. Subsequently, profitability is assessed using profit contribution rates, and profitable vegetable products are selected. Finally, optimal replenishment and pricing strategies are proposed using linear programming. Rational pricing and replenishment strategies not only enhance supermarket operational efficiency but also contribute to the sustainable development of the vegetable supply chain.

Keywords: Seasonal Autoregressive Integrated Moving Average, Multiple Linear Regression, Correlation Analysis, Time Series Analysis

1 Introduction

With the development of the socio-economic landscape and the increasing awareness of public health, vegetables, as an essential food in daily life, have seen a growing demand. This is particularly true in fresh food supermarkets, where vegetables, being perishable, have a relatively short shelf life and experience significant price fluctuations. These characteristics present several challenges for supply chain management. How to optimize inventory management, reduce food waste, and maximize profits through scientific pricing and replenishment strategies has become a key issue in the fresh food industry. Effective vegetable pricing and replenishment strategies not only improve supermarket operational efficiency but also enhance consumer satisfaction, thereby driving the sustainable development of the entire supply chain. Consequently, research on vegetable pricing and replenishment strategies, especially through the application of

W. Zang and C. Xia (eds.), *Proceedings of the 2025 3rd International Conference on Digital Economy and Management Science (CDEMS 2025)*, Advances in Economics, Business and Management Research 336,
https://doi.org/10.2991/978-94-6463-770-0_63

appropriate mathematical modeling and optimization methods, is a crucial topic for decision-making support in pricing and replenishment within supply chain management.

In vegetable supply chain management, numerous scholars have employed various mathematical models and algorithms to address the pricing and replenishment problem. Shen et al. [1] optimized replenishment strategies and predicted price fluctuations by combining linear regression models with the Autoregressive Integrated Moving Average Model time series model, enabling more accurate pricing decisions. Guo and Zhao [2] used a combination of Long Short-Term Memory Networks and Extreme Gradient Boosting) models to forecast vegetable sales and optimize pricing strategies. Lv et al. [3] applied genetic algorithms to optimize the replenishment quantities and pricing strategies for vegetables with the goal of maximizing supermarket revenue. Geng et al. [4] employed the Particle Swarm Optimization algorithm to optimize vegetable pricing and replenishment strategies, taking into account seasonal fluctuations and supply-demand relationships. Tang [5] proposed a joint machine learning model based on LSTM and XGBoost to optimize pricing and replenishment decisions in the vegetable supply chain, thereby enhancing the overall revenue of supermarkets.

This study addresses the pricing and replenishment challenges in the vegetable commodity market by proposing a comprehensive optimization model. Initially, a correlation analysis and time series analysis are conducted on the sales data of various vegetable categories. Based on the identified correlations and seasonal characteristics of vegetable sales[1], a multiple linear regression model and a forecasting model are developed. The multiple linear regression model quantifies the relationship between the pricing and sales of different vegetable categories. Furthermore, the SARIMA model is employed to predict the sales of each vegetable category for the forthcoming week, facilitating precise short-term sales forecasting. Finally, by integrating a linear programming model, optimal replenishment and pricing strategies are proposed to maximize the supermarket's revenue. By utilizing these models and algorithms, this study not only enables scientific demand forecasting for vegetable products but also allows the formulation of rational replenishment and pricing strategies to enhance operational efficiency and profitability in supermarkets.

2 Model Development and Solution

In practical scenarios, the sales of various vegetable categories may exhibit certain substitution and complementarity relationships, which significantly influence pricing and replenishment decisions. To comprehensively account for these factors, this paper constructs a multiple regression model based on the cross-price elasticity of demand. During the model development process, the Shapiro-Wilk (SW) test is first employed to assess the normality of the sales and pricing data for each vegetable category. Following this, a correlation analysis is performed, and appropriate correlation metrics are selected: the Pearson correlation coefficient is applied if the variables follow a normal distribution, while the Spearman rank correlation coefficient is used if the variables do not

[1] This paper is based on the test data of this web page: https://www.mcm.edu.cn/html_cn/node/c74d72127066f510a5723a94b5323a26.html.

conform to normality. Based on the results of the correlation analysis, a regression equation is established to form the pricing model. Additionally, a time series analysis is conducted on the sales data of each vegetable category to predict both the sales and wholesale prices. Using these predictive results, the objective function and constraints of the optimization model are formulated, incorporating shelf display limitations as constraints. Ultimately, the model seeks the optimal solution under the objective of profit maximization.

2.1 Correlation Analysis and Time Series Analysis

A. Correlation Analysis.

The sales of different types of vegetables may influence each other. To explore whether there exists a complementary or substitute relationship, this paper conducts a correlation analysis of the monthly sales volumes for various vegetable categories.

In order to perform an effective statistical analysis of the sales and price data for vegetable categories, it is first necessary to test whether these data follow a normal distribution. For this purpose, the SW test is employed. This test helps determine whether the data conform to the normality assumption, providing a basis for subsequent correlation analysis. The specific steps of the SW test are as follows:

Step 1: Establish the Null Hypothesis and Alternative Hypothesis

Null hypothesis H_0: The data follow a normal distribution.

Alternative hypothesis H_1: The data do not follow a normal distribution.

Step 2: Calculate the W Statistic

W is the statistic used in the SW test, and it is calculated based on the following formula.

$$W = \frac{\left(\sum a_i x_i\right)^2}{\sum\left(x_i - \bar{x}\right)^2}. \tag{1}$$

Where x_i is the i-th data point, $\bar{x}$ is the mean of the data, and a_i are constants related to the sample size, which come from precomputed tables. These coefficients are derived from the expected values and covariance matrix of the normal distribution, with the goal of maximizing the efficiency of the W statistic.

Step 3: Find the Corresponding p-value

The W statistic is the main output of the SW test, and it measures the degree of correspondence between the sample data and a perfect normal distribution. The W value ranges from 0 to 1, where a value closer to 1 indicates that the data are more similar to a normal distribution. The p-value is a commonly used measure in hypothesis testing, representing the probability of observing the current or more extreme test statistics, assuming the null hypothesis is true. After calculating the W value, the corresponding p-value can be obtained using statistical software or lookup tables (Table 1).

Table 1. Normality test results.

	Leafy vegetables	Floral vegetables	Aquatic root and rhizome vegetables	Solanum vegetables	Chili peppers	Edible mushrooms
W statistic	0.713	0.869	0.905	0.971	0.953	0.919
p-value	0.001	0.198	0.250	0.896	0.703	0.349

Step 4: Interpretation of Results

Given the significance level α (typically set at $\alpha = 0.05$), when $p > \alpha$, the null hypothesis is not rejected, implying that the data follow a normal distribution; when $p \leqslant \alpha$, the null hypothesis is rejected, indicating that the data do not follow a normal distribution. In this paper, normality tests were conducted on the existing data. Based on the normality test of monthly sales, the sales of Floral vegetables do not follow a normal distribution, while the sales of the other vegetable categories follow a normal distribution. Therefore, the Spearman rank correlation test was chosen for further analysis.

The Spearman correlation coefficient is given by:

$$\rho = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(R(x_i) - \overline{R(x)}\right)\cdot\left(R(y_i) - \overline{R(y)}\right)}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\left(R(x_i) - \overline{R(x)}\right)^2\right)\cdot\left(\frac{1}{n}\sum_{i=1}^{n}\left(R(y_i) - \overline{R(y)}\right)^2\right)}}. \quad (2)$$

where $R(x)$ and $R(y)$ represent the ranks of x and y, $\overline{R(x)}$ and $\overline{R(y)}$ are the average ranks of x and y, respectively.

The Spearman correlation coefficients were visualized in a heatmap ((Fig. 1), revealing that there is a significant correlation in the sales across different vegetable categories. Notably, the correlation between Edible mushrooms and Aquatic root and rhizome vegetables, Chili peppers, and Solanum vegetables is particularly high, with correlation coefficients of 0.927 and 0.806, respectively.

B. Time Series Analysis.

Vegetables typically exhibit strong seasonality. If the sales of vegetable categories show significant seasonal variation through time series analysis, seasonal factors must be considered in the sales forecasting. To test the seasonality of the sales across different vegetable categories, time series analysis was performed on the monthly sales data, and the results revealed significant seasonal differences in the sales of the vegetable categories (Fig. 2).

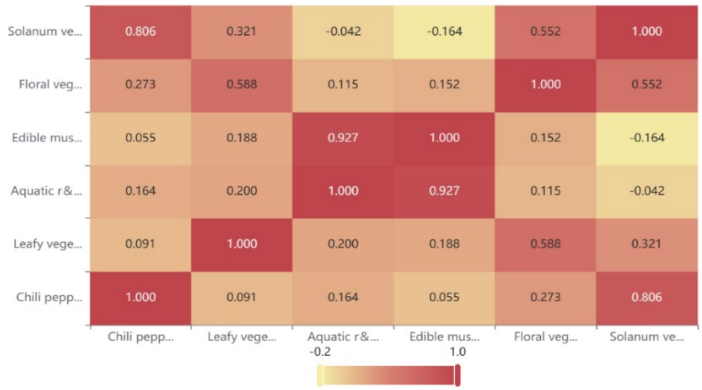

Fig. 1. Heat map of correlation coefficient of sales volume of each vegetable category.

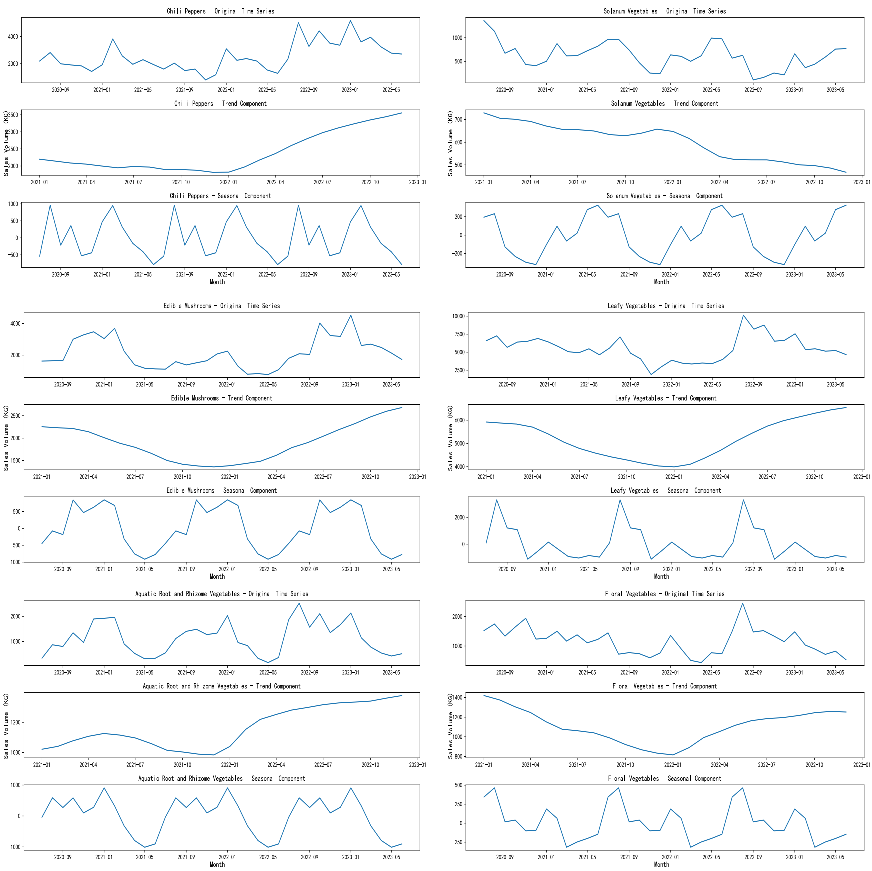

Fig. 2. Time series distribution of monthly sales volume of each vegetable category.

2.2 Forecasting Sales and Wholesale Prices for Vegetable Categories

The results from the time series analysis indicated that the sales of vegetable categories exhibit significant seasonality. Therefore, this paper employs the SARIMA (Seasonal Autoregressive Integrated Moving Average) model to forecast the sales for each vegetable category for the upcoming week (July 1-7, 2023).

The SARIMA model is one of the time series forecasting methods. SARIMA extends the ARIMA model by incorporating seasonal factors, thus overcoming the limitation of ARIMA, which cannot account for cyclical patterns.

The ARIMA model consists of three components: Autoregressive (AR), Moving Average (MA), and Integration (I). The AR part is represented by the parameter p, which describes the influence of previous values in the time series on the current value; the MA part is represented by the parameter q, which captures the relationship between the current error and past errors in the time series; the I part is represented by the parameter d, which transforms a non-stationary time series into a stationary one.

SARIMA also consists of three components: Seasonal Autoregressive (SAR), Seasonal Moving Average (SMA), and Seasonal Integration (SI), represented by the uppercase parameters P, Q, and D, respectively, which introduce seasonal factors. These parameters, along with the seasonal period s, define the SARIMA model as follows:

$$SARIMA(p,d,q) \times (P,D,Q)_s. \tag{3}$$

This study employs the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) diagrams to determine the non-seasonal parameters, and uses the Akaike Information Criterion (AIC) to identify the seasonal parameters.

(1) Identification of Non-Seasonal Factors

To identify non-seasonal factors, this study utilizes the ACF and PACF (Fig. 3). Due to space constraints, the results for Edible mushrooms and Leafy vegetables are presented in this paper. Based on the analysis, the optimal non-seasonal ARIMA model for both Edible mushrooms and Leafy vegetables was determined to be $(0,1,1)$.

(2) Identification of Seasonal Factors

To more accurately capture seasonal fluctuations in the data, this study applies the AIC criterion for selecting the seasonal parameters of the model. A grid search is conducted within a given constraint range, and the optimal seasonal parameters are identified which are shown in Table 2.

Table 2. Seasonal parameter.

Vegetable category	Seasonal parameter	Vegetable category	Seasonal parameter
Leafy vegetables	$(1,1,2)_{12}$	Edible mushrooms	$(1,1,2)_{12}$
Floral vegetables	$(1,1,2)_{12}$	Solanum vegetables	$(1,1,2)_{12}$
Aquatic root and rhizome vegetables	$(2,1,1)_{12}$	Chili peppers	$(1,1,2)_{12}$

By combining the non-seasonal and seasonal factors, the SARIMA model is used to forecast the sales volume of each vegetable category. The results are shown in Fig. 4.

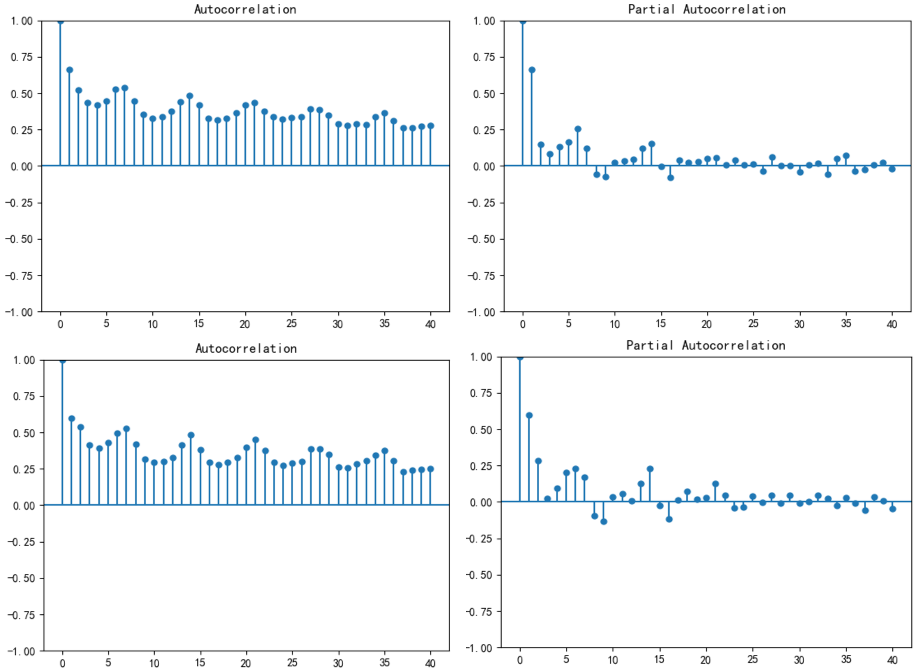

Fig. 3. ACF diagram and PACF diagram partial results are displayed.

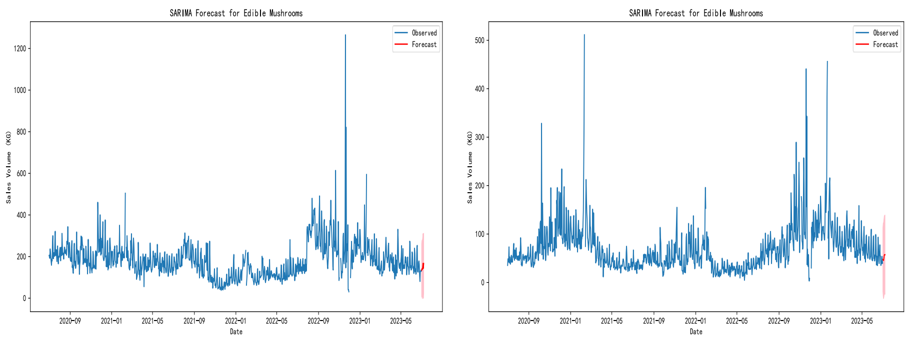

Fig. 4. SARIMA model prediction results.

Based on the predictions from the SARIMA model, the sales trends of different vegetable categories are further analyzed. The forecasted results show that the sales volume trends for the various vegetable categories in the upcoming week are as illustrated in Fig. 5.

Subsequently, the wholesale prices for each vegetable category in the coming week are estimated. Given the inherent seasonal fluctuations in vegetable prices and the fact that vegetables are essential commodities, their prices typically remain relatively stable over the long term. In this study, distribution charts of the wholesale prices for each vegetable category in July of 2022, 2021, and 2020 are plotted to observe the price variation patterns.

It can be observed in Fig. 6, in general, the wholesale prices of vegetables show minimal fluctuation during the same period each year. Therefore, it is reasonable to assume that the wholesale prices of vegetables in July 2023 will not differ significantly from those of the same period in the previous three years. As a result, the average wholesale prices for July over the past three years are used to estimate this year's prices. This estimation method, which only considers data for the same month, inherently controls for seasonal factors, making it a simple yet effective approach.

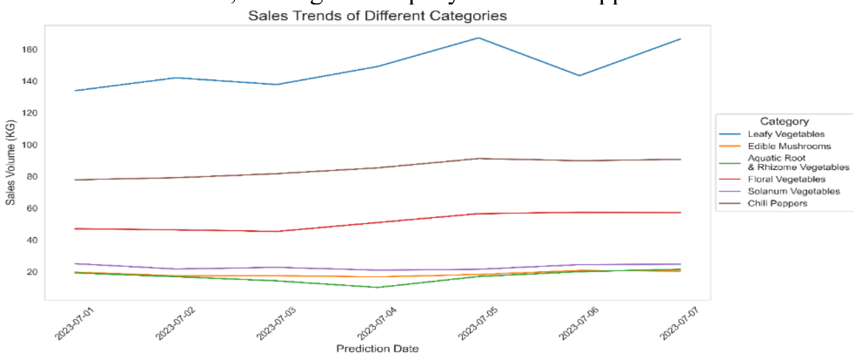

Fig. 5. Forecast of the next week's sales volume of each vegetable category.

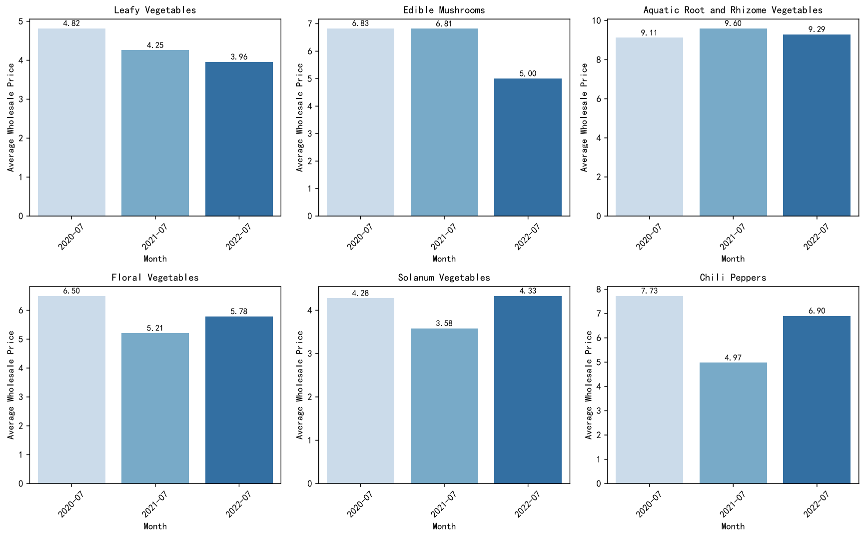

Fig. 6. Comparison of wholesale prices of vegetables in July over the past year.

2.3 Establishing a Vegetable Pricing Model

After completing the correlation analysis of the sales data, this study constructs a multiple regression model based on cross-price elasticity to quantitatively describe the interrelationship between vegetable pricing and sales volume. This model better captures the impact of price changes in one vegetable category on the sales of other categories, thereby optimizing pricing and replenishment strategies.

$$Q_i = \sum_{j=1}^{6} \eta_{ij} P_j + \lambda + \epsilon. \tag{4}$$

$$\eta_{ij} = \frac{\mathrm{d}Q_i/Q_i}{\mathrm{d}P_j/P_j}. \tag{5}$$

In the equation, Q_i represents the sales volume of the i-th vegetable category, n_{ij} is the cross-price elasticity between the i-th and j-th vegetable categories, which measures the change in the sales volume of the i-th vegetable category when the price of the j-th vegetable category increases by one unit. P_j represents the price of the j-th vegetable category, and i and j correspond to the vegetable categories, ranging from 1 to 6: Leafy vegetables, Floral vegetables, Aquatic root and rhizome vegetables, Edible mushrooms, Solanum vegetables, and Chili peppers. λ represents the intercept of the equation, and ϵ denotes the error term.

The price of each individual product (w_i) is determined by the profit margin (r_i) of the respective vegetable category.

$$r_i = \frac{P_i - C_i}{C_i} \tag{6}$$

$$w_i = (1 + r_i) \times c_i \tag{7}$$

In the equation, C_i is the wholesale price of the i-th vegetable category, c_i is the wholesale price of the i-th individual vegetable product, and w_i is the pricing of the i-th individual vegetable product.

2.4 Determination of the Linear Programming Model

In practical sales scenarios, shelf space is limited. To control the total number of individual items between 27 and 33, this study employs the profit contribution rate as a metric to quantify the profitability of individual vegetable products. The products are ranked based on their profit contribution rates, and those with the highest rankings are selected for sale. The profit contribution rate is defined as the percentage of the total supermarket vegetable sales profit generated by the individual vegetable product over the past seven days.

Step 1: Constructing the Objective Function

$$Y = \sum_{i=1}^{n} (p_i - c_i) \times q_i. \tag{8}$$

The objective function is formulated where q_i represents the sales volume of the i-th vegetable product.

Step 2: Determining the Constraints

$$27 < i < 33, N_i > 2.5, N_i \geq q_i. \tag{9}$$

This constraint ensures that the replenishment volume does not fall below the minimum market demand nor exceed the supermarket's storage capacity.

2.5 Formulation of Pricing and Replenishment Strategies

Based on the model constructed in this study, pricing and replenishment strategies for 27 individual vegetable products are derived. A selection of these results is presented in Table 3.

Table 3. Pricing and replenishment strategies.

Product	Price	Replenishment Quantity (kg)	Product	Price	Replenishment Quantity (kg)
Sweet Potato Shoots	5.7	9	Baby Chinese Cabbage	8.19	14
Spinach	7.24	8	Shanghai Qing	6.87	6
Yunnan Romaine Lettuce	4.3	17	Broccoli	9.77	14
Small Chili Peppers	9.12	27	Amaranth	3.99	10
Wood Ear Vegetable	4.13	7	Baby Napa Cabbage	6.11	14

3 Model Evaluation and Promotion

Through the above model, the pricing and replenishment strategies of the supermarket are well formulated, but there are still some places that can be modified and improved:

(1) Based on the vegetable sales context in supermarkets, this study establishes the SARIMA model, specifically designed for time series data, which provides both large-scale and small-scale time series forecasts. With this model, supermarkets can not only predict long-term sales trends but also achieve precise short-term forecasts for specific dates or time periods.

(2) While the multiple linear regression model addresses the correlation between supply and demand factors, it remains fundamentally a linear approach. In reality, the

relationship between supply and demand may be nonlinear and dynamic, influenced by multiple interacting factors.

4 Conclusion

This study presents a novel method to tackle pricing and replenishment issues within the vegetable supply chain, particularly for fresh food supermarkets. By integrating correlation analysis, time series forecasting, and optimization techniques, it proposes an effective approach to enhance inventory management, boost supermarket profits, and minimize food waste.

We suggest a multiple linear regression model to quantify the connection between prices and sales, alongside a SARIMA forecasting model to manage seasonal variations in vegetable demand. These models facilitate accurate predictions of short-term sales, allowing supermarkets to modify their replenishment strategies in real-time. Furthermore, by incorporating linear programming, we can identify optimal prices and replenishment quantities, thus ensuring efficient resource allocation grounded in profitability.

A significant contribution of this research lies in the creation of a pricing strategy that utilizes cross-price elasticity to acknowledge the interdependence of various vegetable categories. This model not only enhances supermarket revenue but also promotes a more adaptable and responsive reaction to market changes. Additionally, the optimization model considers shelf space limitations, which is vital for supermarkets dealing with perishable items.

While the proposed models serve as valuable decision-making tools, we recognize certain limitations, especially regarding the assumption of linear relationships among variables. Future research could refine these models by exploring nonlinear dynamics and other intricate factors affecting vegetable supply and demand.

In summary, this study significantly advances vegetable supply chain management by presenting actionable strategies for pricing and replenishment, ultimately improving both supermarket profitability and sustainability in the vegetable sector.

Acknowledgments

This work was supported by the Natural Science Foundation of Top Talent of SZTU (grant no. GDRC202333).

References

1. Shen, Y., Zhao, Y., & Hao, X. (2024). Automated Pricing and Replenishment Decisions for Vegetable Products Based on Data Analysis. Highlights in Science, Engineering and Technology.
2. Guo, X., & Zhao, G. (2023). Automatic Pricing and Replenishment Strategy for Vegetable Products Based on XGBoost and LSTM Prediction Models. Highlights in Science, Engineering and Technology.

3. Lv, Y., Sun, X., & Liu, Q. (2024). Research on optimal pricing and replenishment strategy of vegetable commodities. Highlights in Business, Economics and Management.
4. Geng, J., Zhai, Y., & Yao, H. (2023). Research on vegetable commodity pricing replenishment decision based on particle swarm algorithm. 2023 IEEE International Conference on Electrical, Automation and Computer Engineering (ICEACE), 1069-1073.
5. Tang, Y. (2024). Intelligent Optimization Analysis of Automatic Pricing and Replenishment of Vegetable Products Based on Combinatorial Machine Learning Models. 2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), 511-516.