Transactions on Economics, Business and Management Research
ISSN: 2960-1789, eISSN: 2960-2254 | Volume 3
MSIE 2023

The Research on Pricing and Replenishment Optimization of Fresh Supermarket Vegetable Products based on Sales Data

Jintao Yang *, Zhuo Liu, Shuo Huang, Gangtong Zhou, Weizhi Li
Northeast Forestry University, College of Civil and Transportation Engineering, Harbin, China
* Corresponding Author Email: 2642689088@qq.com

Abstract. This study delves into the challenges and complexities of vegetable sales management in modern commercial environments, particularly in fresh food supermarkets. By utilizing descriptive statistics and visual analysis of sales data, the study reveals the correlations between different vegetable categories and quantifies these correlations using Pearson's correlation coefficient. Subsequently, by integrating time series analysis and multi-objective programming, a mathematical model is constructed, aimed at maximizing profits under specific constraints. The innovation of this research lies in its comprehensive consideration of category-level sales management and the application of modern optimization algorithms for replenishment and pricing strategies. The uniqueness of this paper is in its integrative approach to the vegetable sales problem, providing refined mathematical models and advanced optimization methods. Finally, the paper thoroughly describes the steps of the research design, including sales data analysis, cost markup analysis, and the construction of an optimization model based on time series analysis and multi-objective programming, intending to offer supermarkets a vegetable sales management plan adaptable to market changes.

Keywords: Vegetable Sales Management; Supermarket Replenishment and Pricing Decisions; Sales Trend Correlation Analysis; Time Series Forecasting Analysis; Multi-objective Programming Model.

1. Introduction

In today's rapidly evolving business environment, vegetable sales, as a core component of the food supply chain, face multifaceted and complex challenges. With the increase in consumer spending levels and the shift in consumer mindsets, how to formulate pricing and replenishment strategies has become an urgent issue for fresh supermarkets to address[1]. Supermarkets typically need to make daily replenishment decisions based on the historical sales and demand of products. However, the procurement of vegetables usually takes place in the early hours, between 3:00 to 4:00 AM. Without clear information on specific items and purchase prices, supermarkets are required to make replenishment decisions for various vegetable categories within a very short time. Therefore, in this context, deeply mining sales trends and correlations, while comprehensively considering the balance between supply and demand, becomes key to optimizing supermarket operations.

To accurately formulate pricing and replenishment strategies for vegetable products, Fan (2019) applied regression models to price vegetables[2]. Zou et al. (2018) proposed a joint replenishment method for multiple perishable products with stochastic demand[3].

In this study, methods involving data analysis, mathematical modeling, and optimization algorithms are comprehensively utilized. Initially, descriptive statistical analysis and visual analysis are conducted on sales data, cost per unit, and loss rate, to reveal the sales correlations between different vegetable categories. For this, we employed correlation analysis, specifically using the Pearson correlation coefficient to quantify the degree of association between different vegetable categories. Subsequently, we applied mathematical modeling to develop a system based on time series analysis and a combination of multi-objective programming. The aim of this model is to maximize supermarket profits while satisfying various constraints.

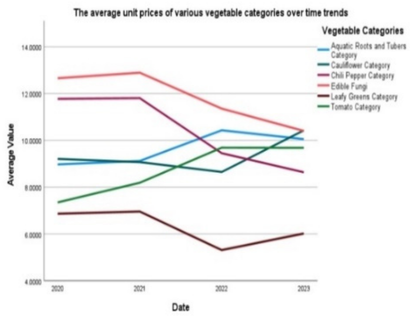
The innovation of this study lies in its comprehensive consideration of category-level sales management. By conducting an in-depth analysis of sales data and constructing an optimized model, and utilizing modern optimization algorithms, it provides a practical and feasible decision support system for supermarkets. Through scientifically rigorous mathematical modeling and the application of modern optimization algorithms, we aim to provide supermarkets with concrete and feasible category-level replenishment and pricing strategies. Retailers can reduce losses and increase profits through pricing and replenishment strategies[4]. To adapt to the constantly changing market environment and ultimately achieve maximum economic benefits, we aim to optimize vegetable sales management. This will empower supermarkets, enhancing their competitiveness in the fierce market competition, and simultaneously make a positive contribution to the sustainable development of the entire food supply chain.

2. Data Selection and Visualization Analysis

2.1. Data Acquisition

In the context of this study, a specific supermarket chain was selected as the core subject of the research. The dataset encompasses comprehensive product information for six different categories of vegetables offered in the market. It includes complex transaction records, encompassing sales details and wholesale price data, as well as the spoilage rates of six vegetable categories, spanning from July 1, 2020, to June 30, 2023.

2.2. Data Visualization Analysis



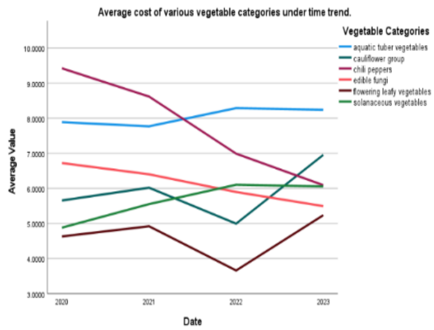


Figure 1. Average unit prices of various vegetable **Figure 2.** Average cost of various vegetable

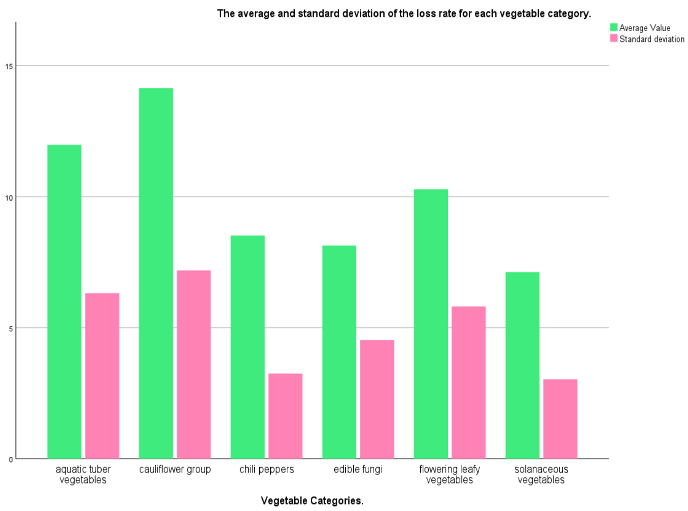


Figure 3. The average and standard deviation of the loss rate for each vegetable category

The overall observation of the historical changes in the sales and cost prices of the six vegetable categories over the past few years reveals different trends and fluctuations. Specifically, the sales price of cruciferous vegetables fluctuated significantly from 2020 to 2023, with a notable increase between 2022 and 2023; leafy vegetables, on the other hand, showed a downward trend. The price of eggplant-type vegetables was relatively stable, while chili-type vegetables have seen considerable fluctuations in recent years, especially with a significant decrease between 2022 and 2023. The price of edible mushrooms varied significantly, and aquatic root and stem vegetables saw a year-over-year increase, particularly with a noticeable rise between 2021 and 2022. In terms of cost prices, cruciferous vegetables and aquatic root and stem vegetables exhibited a fluctuating upward trend, while leafy vegetables and edible mushrooms showed a downward trend. The cost price of eggplant-type vegetables generally increased, whereas chili-type prices declined from 2020 to 2022, with a slight rebound in 2023. These changes are likely influenced by factors such as market demand, production costs, and supply chain issues.

The analysis of the spoilage rates for the six vegetable categories shows that cruciferous vegetables have the highest spoilage rate at 14.14%, followed by aquatic root and stem vegetables and leafy vegetables at 11.97% and 10.28%, respectively. Eggplant-type vegetables, chili-type vegetables, and edible mushrooms have relatively lower spoilage rates, at 7.12%, 8.52%, and 8.13% respectively. In terms of volatility, the spoilage rates of cruciferous vegetables and aquatic root and stem vegetables fluctuate more, while those of eggplant-type and chili-type vegetables are more stable. This indicates that cruciferous vegetables, aquatic root and stem vegetables, and leafy vegetables are more susceptible to spoilage, while the other categories are relatively stable.

3. Methodology

3.1. Research Design

To thoroughly address the issues of vegetable sales management in fresh food supermarkets, this study emphasizes the real-life context of fresh food supermarket vegetable sales management in its design. Firstly, descriptive statistical analysis and visualization analysis are conducted using sales data and relevant information. Additionally, correlation analysis methods are employed to deeply study the sales data, revealing the sales interconnections among different vegetable categories. This ensures the reliability of subsequent analysis and modeling.

Cost mark-up analysis is another key step in the design of this study. By conducting a detailed analysis of the cost, retail price, and wholesale price of each vegetable category, and calculating the cost mark-up ratio, we gain a deeper understanding of the relationship between total sales volume and cost mark-up. This assists in establishing a more precise optimization model, while also taking into account the economic benefits of the supermarket.

During the optimization model construction phase, we will design a model based on a combination of time series analysis and multi-objective planning, grounded in the results of sales trend and cost mark-up analysis. Methods typically used in sales forecasting include time series algorithms[5], The ARIMA model, a time series analysis method, will be used to predict the future one-month sales unit price, sales volume, and cost unit price. The multi-objective planning model will consider multiple constraints, such as storage limits and financial budget constraints, to ensure feasibility in practical operations. The ultimate optimization goal of the model will be defined as maximizing the total revenue of the supermarket in the upcoming month.

This approach aims to provide more accurate demand forecasting for supermarkets. Accurate demand forecasting is crucial for any small to medium-sized manufacturer, as it significantly aids in precise inventory management, cost control, and strategic planning, ensuring efficiency and profitability[6].

3.2. Model Establishment and Solution

3.2.1. Pearson Correlation Coefficient

The Pearson Correlation Coefficient is a statistical measure used to quantify the strength and direction of the linear relationship between two continuous numerical variables. The coefficient, denoted as r describes the degree of linear correlation between two variables, with the absolute value of r indicating the strength of the linear relationship. The larger the absolute value of r, the stronger the linear correlation between the two variables. The Pearson Correlation Coefficient is employed to measure the strength and direction of the linear relationship between two continuous variables.

The Pearson correlation coefficient formula is:

$$r = \frac{\Sigma_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma_{i=1}^{n}(x_i - \bar{x})\sqrt{\Sigma_{i=1}^{n}(y_i - \bar{y})}^2}} \tag{1}$$

The range of this coefficient is from -1 to +1, where -1 indicates a perfect negative linear relationship, +1 indicates a perfect positive linear relationship, and 0 means there is no linear relationship.

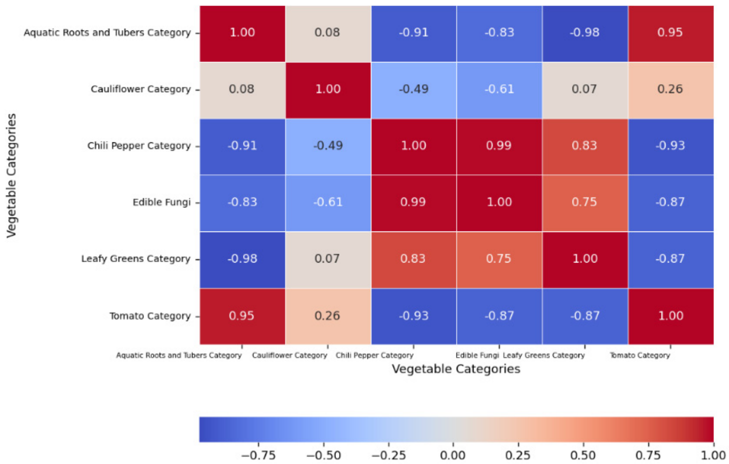


Figure 4. Pearson correlation coefficient heatmap

Observing the heatmap, we can see some distinct trends. There is a higher correlation between the "Cauliflower Category" and the "Leafy Greens Category," showing a positive relationship. In contrast, the correlation between the "Leafy Greens Category" and the "Chili Pepper Category" is lower, indicating a weaker linear relationship.

Additionally, there is a positive correlation between the "Tomato Category" and the "Cauliflower Category," while its correlation with both the "Leafy Greens Category" and the "Chili Pepper Category" is relatively lower.

3.3. Calculate the Cost Markup Ratio

For each vegetable category, calculate its cost markup ratio. The cost markup ratio can be calculated using the following formula:

$$Cost\ Markup\ Ratio\ (\%) = \frac{\text{Retail Price per Unit}}{\text{Wholesale Price per Unit}} \times 100\% \tag{2}$$

The analysis of the cost markup rates for six categories of vegetables in this study reveals significant market volatility and pricing differences. The data indicates that aquatic roots and tuber vegetables have the highest markup rates, while the pricing of chili pepper varieties is relatively stable. Stacked bar charts visually present the averages, medians, and extremes between different categories, highlighting the market's diversity and potential profit opportunities.

Table 1. Average, median, maximum, and minimum of cost markup rates

Vegetable categories	Average	Group median	Maximum value	Minimum value
Aquatic root and	1904.86	569.33	6118.09	362.69
Chili pepper varieties	1313.36	1236.41	2275.89	504.73
Edible mushrooms	1398.03	1230.28	2292.45	839.11
Eggplant varieties	146.41	144.30	235.03	62.00
Flowering and leafy	871.51	790.86	1360.75	543.57
Flowering vegetables	473.53	406.99	958.37	121.78

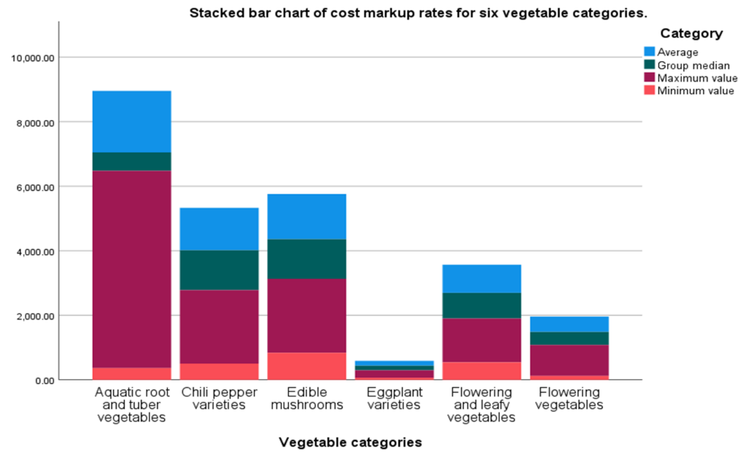
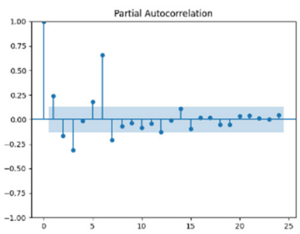


Figure 5. Stacked bar chart of cost markup rates for six vegetable categories

3.4. Time Series Analysis

Classic time series forecasting methods include Autoregressive models (AR), Moving Average models (MA), and Autoregressive Integrated Moving Average models (ARIMA)[7], A time series is a sequence of values arranged in chronological order, also known as a dynamic series[8]. Time series forecasting is a widely used method in continuous data sequences[9], Accurate time series forecasting is crucial, especially in practical application areas such as demand forecasting[10].



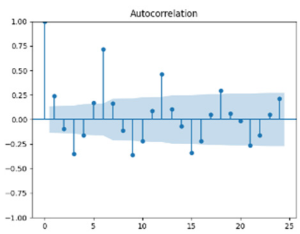


Figure 6. PACFplot **Figure 7.** ACFplot

First, conduct a time series analysis of data from June 2020 to June 2023 using SPSS software, applying the ARIMA model (which combines autoregression, differencing, and moving average) to forecast future prices. The ADF test result shows a statistic of -3.461354 and a p-value of 0.009037, indicating that the data is stationary at the 5% significance level, without the need for differencing, Analysis of the PACF and ACF plots suggests preliminary ARIMA model parameters as p=1, d=0, q=1, which corresponds to the ARIMA(1,0,1) model.

Table 2. ARIMA model fitting results table

Statistical fitting	Mean	Minimum	Maximum	Percentile value						
				5	10	20	50	75	90	95
R^2	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988
RMSE	0.105	0.105	0.105	0.105	0.105	0.105	0.105	0.105	0.105	0.105
MAPE	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
MaxAPE	0.043	0.043	0.043	0.043	0.043	0.043	0.043	0.043	0.043	0.043
MAE	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
MaxAE	0.876	0.876	0.876	0.876	0.876	0.876	0.876	0.876	0.876	0.876
Normalization.BIC	-4.492	-4.492	-4.492	-4.492	-4.492	-4.492	-4.492	-4.492	-4.492	-4.492

According to the fitting results of the ARIMA model in the table, the model exhibits a series of significant advantages. First, the R-squared value is as high as 0.988, meaning the model can explain 98.8% of the variance in the data, indicating a very high explanatory power. Second, the Root Mean Square Error (RMSE) is only 0.105, suggesting that the deviation between the model's predictions and the actual observations is relatively small. Additionally, the Mean Absolute Percentage Error (MAPE) value of 0.001 shows extremely high accuracy in prediction. Moreover, the Mean Absolute Error (MAE) value of only 0.024 further confirms the model's high precision in handling data. Lastly, the negative value of the Bayesian Information Criterion (BIC) (-4.492) indicates a very good fit, considering the complexity of the model. These statistics collectively point to a conclusion: the ARIMA model excels in both predictive capability and model fitting.

Using the ARIMA model for time prediction, the forecast of the sales unit price, cost unit price, and sales volume for six categories of vegetables for the upcoming month is as Table 3:

Table 3. Pricing and procurement strategy for the supermarket for the next month.

Category	Cost unit price	Unit selling price	Quantity of sale
Edible Fungi	6.66Yuan	14.39Yuan	1620.10kg
Chili Pepper Category	7.18Yuan	9.27Yuan	2487.20kg
Tomato Category	4.48Yuan	7.57Yuan	1258.22kg
Leafy Greens Category	5.08Yuan	10.56Yuan	7723.26kg
Cauliflower Category	6.34Yuan	7.81Yuan	1962.29kg
Aquatic Roots and Tubers Category	9.89Yuan	11.04Yuan	729.66kg

3.5. Multi-objective Programming Model

To better formulate the replenishment and pricing decisions for vegetable products, this paper establishes a multi-objective programming model aimed at solving for the maximization of the supermarket's total profit. This model takes into account the retail price, wholesale price, sales volume, and loss rate of the products, forming an objective function. The formula is as follows:

$$ProfitM = \Sigma_{i=1}^{n}(Pi \times Qi - Ci \times Qi - Oi) \tag{3}$$

Constraint 1: Inventory Constraint.

$$I_i \leqslant I_{\max} \tag{4}$$

Constraint 2: Demand Constraint.

$$S_i \leqslant D_i \tag{5}$$

Constraint 3: Financial Budget Constraint.

$$\Sigma_i C_i \times Q_i \leqslant B \tag{6}$$

Constraint 4: Constraint on the Relationship between Retail Price and Wholesale Price.

$$P_i \geqslant C_i \tag{7}$$

Constraint 5: Loss Rate Constraint.

$$O_i \leqslant E_i \times W_i \tag{8}$$

In summary, the multi-objective programming model is established as follows:

$$ProfitM = \Sigma_{i=1}^{n}(Pi \times Qi - Ci \times Qi - Oi)$$

$$\left\{ \begin{array}{c} I_i \leqslant I_{\max} \\ S_i \leqslant D_i \\ \Sigma_i C_i \times Q_i \leqslant B \\ P_i \geqslant C_i \\ O_i \leqslant E_i \times W_i \end{array} \right\} \tag{9}$$

Based on the forecasted data, this function calculates a total profit of approximately 58,200.38 yuan. The model comprehensively considers various factors, including the retail price, wholesale price, sales volume, and loss rate of the products. In the specific calculation process, the sales revenue for each category of vegetables is first determined, which is the product of the unit sales price and sales volume. Next, the cost of each vegetable is calculated by multiplying the cost per unit with the sales volume. Then, considering the impact of the loss rate, the cost incurred due to loss is further calculated by multiplying the aforementioned costs with their respective loss rates. Finally, the total profit is obtained by summing up the sales revenue of all vegetables and subtracting all costs and loss-induced costs.

4. Discussion

In the rapidly evolving commercial environment, vegetable sales, as a core component of the food supply chain, face multiple complex challenges. This research, by delving into the replenishment and pricing decisions of vegetable commodities in fresh food supermarkets, highlights the criticality of these decisions in supermarket management. Adopting a comprehensive approach of data analysis, mathematical modeling, and optimization algorithms, the study conducts descriptive statistical and visual analyses of sales data, revealing the sales correlations between different vegetable categories. By utilizing the Pearson correlation coefficient and a time series-based linear programming model, this research not only proposes new strategies for optimizing supermarket operations but also demonstrates an innovative approach to considering category-level sales management comprehensively, providing a practical decision-support system for supermarkets.

5. Conclusion

This thesis conducts an in-depth analysis of the complexity of vegetable sales management in fresh supermarkets, with a special focus on the constantly changing and uncertain commercial environment.

The study uncovers significant sales correlations among different vegetable categories, quantifying these using the Pearson correlation coefficient. This approach breaks through the limitations of traditional analysis, demonstrating the innovation and practicality of data analysis. By integrating time series analysis and multi-objective programming, the thesis constructs a mathematical model that considers not only sales data and costs but also factors like loss rates, thereby optimizing supermarket profits and operational efficiency comprehensively. Additionally, the thesis emphasizes the importance of an interdisciplinary approach, combining data analysis, mathematical modeling, and modern optimization algorithms, showcasing their synergistic effect in business decision-making. The research not only enhances the economic benefits for supermarkets but also provides robust support in the competitive market, contributing new perspectives and methodologies to food supply chain management and sustainable development in the retail industry.

References

[1] Chen T, Zhou L, Guo S. Research on Pricing and Replenishment Strategy for Vegetable Commodities based on Non-linear Programming[J]. Transactions on Economics, Business and Management Research, 2023, p. 19-24.

[2] Fan Fengyan: Research on Supermarket Fresh Food Pricing Model (Master, Fujian Normal University, China 2019) [D], p. 9.

[3] Zou Guangjun, Fu Shaochuan. Joint Replenishment Optimization of Multiple Perishable Products under Random Demand [J]. Management Science and Engineering. Vol. 7 (2018) No.1, p. 9-19.

[4] Zhanyu S X H. Optimal pricing and replenishment policy for perishable food supply chain under inflation[J]. Computers Industrial Engineering,2021,158107433-107433, p. 2-19.

[5] Doganis, P., et al., Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. Journal of Food Engineering, 2006. 75(2), p. 196-204.

[6] Wahedi H, Wrona K, Heltoft M, et al. Improving Accuracy of Time Series Forecasting by Applying an ARIMA-ANN Hybrid Model[C]//IFIP International Conference on Advances in Production Management Systems. Cham: Springer Nature Switzerland, 2022: p. 3-10.

[7] Aram A H A. A novel hybrid model to forecast seasonal and chaotic time series [J]. Expert Systems with Applications, 2024,239p. 2-15.

[8] Fengli J X Z J X R W. Multivariable time series forecasting using model fusion[J]. Information Sciences, 2022, 585262-274. p. 5.

[9] Junior P R, Salomon F L R, de Oliveira Pamplona E. ARIMA: An applied time series forecasting model for the Bovespa stock index[J]. Applied Mathematics, 2014, p. 3.

[10] Yugal P R K. Multi-step time series analysis and forecasting strategy using ARIMA and evolutionary algorithms [J]. International Journal of Information Technology,2021,14(1):p. 1-15.