

Research on Vegetable Replenishment and Pricing Optimization Decision Based on SARIMA Model and Regression Analysis

Peixu Zeng*, Jianheng Zhang

College of Artificial Intelligence, Nanchang Jiaotong Institute, Yichun, China, 330600

* Corresponding Author Email: xuanyuan261624@outlook.com

Abstract. In the context of fresh supermarkets, which are confronted with the challenges of short shelf life and susceptibility to deterioration of vegetable products, these establishments require daily replenishment and careful pricing. To develop more rational pricing strategies, this study starts by using regression analysis to create a linear regression equation that correlates total sales volume with cost-plus pricing. Through a rigorous statistical analysis, a regression equation with heightened reliability is derived to better forecast the relationship between total sales volume and cost-plus pricing. Subsequently, a Seasonal Autoregressive Integrated Moving Average (SARIMA) model is implemented. Drawing on historical sales data, the study forecasts the daily replenishment quantities and pricing strategies for various categories of vegetables in the upcoming week.

Keywords: Linear Regression Equation, SARIMA Model, Pricing Decision.

1. Introduction

In fresh supermarkets, the limited shelf life and the deterioration of quality over time as sales increase present challenges for vegetable products. In general, vegetable varieties that remain unsold on a given day are frequently rendered unsellable the following day. Hence, supermarkets typically require replenishment according to historical sales data and current demand on a daily basis. Given the diverse variety and origin of vegetables, as well as the procurement transactions taking place between 3:00-4:00 in the morning, retailers are required to develop daily replenishment plans for different vegetable categories in the event of unavailable items and uncertain procurement prices. Vegetables typically employ a "cost-plus pricing" approach, offering discounted sales for products of differing quality. Conducting an analysis of market demand is essential for making informed decisions regarding replenishment and pricing. There exists a correlation between the sales volume of vegetables and time from the perspective of demand. The period from April to October is characterized by a wider range of vegetable supplies from a supply perspective. However, the limited selling space in supermarkets underscores the significance of rational sales combinations.

In order to develop more rational pricing strategies, it is crucial to take into account the influence of returns, inventory, and the inherent value of the products on decision-making. Zhao Lin et al. [1] proposed a replenishment and pricing strategy that takes into account returns and fixed costs using stochastic dynamic programming. Meanwhile, Chen Jun et al. [2] examined pricing and inventory replenishment decisions for agricultural products with dual-channel sales. Liu Yizheng et al. [3] conducted a study on a joint replenishment and distribution model for fresh products utilizing RFID technology and its impact on investment decision-making. Cui Ligang et al. [4] employed an adaptive differential evolution algorithm to make replenishment and pricing decisions. Liu Rui [5] investigated approaches for the combined replenishment and distribution of multiple products. Ai Xueyi et al. [6] introduced a joint replenishment strategy for non-instantaneously perishable items with demand that is dependent on deterioration over time. Wang Qiang [7] examined inventory and pricing strategies involving dual replenishment modes and reference price efficiency. Ben Xiangming et al. [8] concentrated on the intelligent implementation of an automated replenishment system. Liu Yang [9] conducted a study that employed the XGBoost algorithm to investigate demand forecasting and replenishment strategies. Le Yi [10] proposed a joint replenishment and distribution strategy utilizing a metaheuristic algorithm.

The Seasonal Autoregressive Integrated Moving Average model (SARIMA) establishes a relationship between replenishment strategies and time, allowing for the replenishment of product combinations with associated temporal patterns. The aim is to enhance revenue, reduce excess inventory, and consequently lower overall expenses.

This paper introduces a procurement and pricing strategy for the forthcoming week in supermarkets, with a specific focus on examining the relationship between sales volume and cost-plus pricing of different categories of vegetables. The interdependence of sales revenue and costs on pricing and procurement quantities underscores the importance of developing suitable functions to depict these relationships. Moreover, the paper anticipates potential replenishment scenarios by considering seasonal factors. The modeling and solution are conducted under the assumption that all goods purchased daily are sold and remain unaffected by factors other than time.

2. Model construction and solution

2.1. Problem description and symbolic explanation

Table 1. Symbol Explanation

Symbol	Explanation
i	Vegetable single-item coding
j	Procurement of vegetables for the j -th occurrence
b_i	The unit selling price of vegetable with item single-item coding i
c_i	The cost of vegetable with item single-item coding i
d_i	The markup rate for vegetable with item single-item coding i
e_{ij}	The procurement of vegetable with item single-item coding i for the j -th occurrence.
f_i	The unit cost of vegetable with item single-item coding i
g_i	The procurement quantity of vegetable with item single-item coding i
h_i	The loss rate for vegetable with item single-item coding i
j_i	The unit fixed cost of vegetable with item single-item coding i
k_i	The unit variable cost of vegetable with item single-item coding i
w_i	The pricing of vegetable with item single-item coding i
l_i	The total profit for vegetable with item single-item coding i

2.2. Model construction

2.2.1 Linear regression equation

The regression equation serves as a mathematical model employed to elucidate the association between an independent variable (or predictor variable) and a dependent variable. The general form of the regression equation can be expressed as follows:

$$Y = A + BX + \epsilon \quad (1)$$

Y -- Dependent variable

X -- Independent variable

A -- Intercept

B -- Regression coefficient

ϵ -- Error term

The aim of regression analysis is to ascertain the most suitable values of A and B in the regression equation, in order to achieve the best fit with the data points and minimize the error term. This is commonly accomplished using statistical techniques, such as the least squares method, with the goal of minimizing the disparities between the predicted and observed values of the dependent variable.

2.2.2 Seasonal Autoregressive Integrated Moving Average model

Seasonal Autoregressive Integrated Moving Average (SARIMA) is a statistical model that has been specifically developed for the purpose of analyzing time series data and making forecasts. The SARIMA model is designed to forecast future values by utilizing historical observations, thereby offering essential assistance for decision-making and strategic planning. This model demonstrates exceptional performance in analyzing time series data characterized by distinct seasonal variations, uncovering trends, seasonal shifts, periodic fluctuations, and other significant time-related attributes. The modeling process consists of the following steps:

(1) The Seasonal Autoregressive Integrated Moving Average (SARIMA) model was trained using historical sales data from a local convenience store, which included total sales and dates.

(2) Utilizing the regression equation previously acquired, forecast the total sales for each vegetable category in the upcoming week while taking into account various pricing strategies. During this process, a range of potential pricing schemes can be developed and applied separately to the time series model in order to generate sales forecasts for each scenario.

(3) To optimize the supermarket's revenue, it is necessary to identify the sales and pricing scenarios that yield the highest revenue for each category by simulating various situations. This may entail the utilization of penalty functions, such as penalty terms incorporated into the loss function, to equilibrate the influence of total sales and pricing.

(4) Finally, considering the optimal total sales and pricing strategy for each category, develop a restocking plan for the upcoming week to maximize the supermarket's revenue.

The specific formulas for time series forecasting models may vary depending on the model type. The Autoregressive Integrated Moving Average (ARIMA) model is widely used in time series prediction. This study chooses to utilize an extended version of the ARIMA model, specifically the Seasonal ARIMA (SARIMA) model.

SARIMA(p,d,q)(P,D,Q,s) Model

p : Autoregressive order, denoting the relationship between the observed values at the current time point and those at the preceding p time points.

d : Differencing order, representing the number of differencing operations required to make a time series stationary. Differencing is performed to eliminate trends and seasonality from the series.

q : Moving average order, indicating the relationship between the observed values at the current time point and the forecast errors from the previous q time points.

P : Seasonal autoregressive order. analogous to p , but applied to the seasonal component.

D : Seasonal differencing order. analogous to d , but applied to the seasonal component.

Q : Seasonal moving average order, analogous to q , but applied to the seasonal component.

s : Seasonal period, the interval during which seasonal effects repeat in the data.

The forecasting formula for the Seasonal Autoregressive Integrated Moving Average (SARIMA) model is as follows:

$$Y_t = c + \varphi_1(Y_{t-1} - c) + \cdots + \varphi_p(Y_{t-p} - c) + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} + \mu_1(Y_{t-s} - c) + \cdots + \mu_P(Y_{t-sP} - c) + \omega_1\varepsilon_{t-s} + \cdots + \omega_Q\varepsilon_{t-sQ} + \varepsilon_t \quad (2)$$

Y_t -- Observed value at time point t

c -- Constant term

φ_p -- Autoregressive coefficient

θ_q -- Moving average coefficient

ε_{t-q} -- Prediction error at time $t-q$

p -- Autoregressive order

q -- Moving average order

μ_P -- Seasonal autoregressive coefficient.

ω_Q -- Seasonal moving average coefficient.

P -- Seasonal autoregressive order

Q -- Seasonal moving average order
s -- Seasonal period

3. Model solution and results analysis

3.1. Analyzing the relationship between the total sales volume and the cost markup pricing for each vegetable category

This study aims to investigate the correlation between the overall sales volume and the cost markup pricing for each category of vegetables using regression analysis. During the process of solving the problem, it was noted that a simple linear regression equation outperformed a multiple linear regression equation. Consequently, we utilize a basic linear regression equation in combination with the cost markup pricing formula to tackle this issue. The cost markup pricing formula can be expressed as follows:

Price = Unit Cost * (1 + Markup Rate)

$$w_i = f_i(1 + d_i) \quad (3)$$

Unit Cost = (Unit Fixed Costs + Unit Variable Costs) / (1 - Loss Rate)

$$f_i = (j_i + k_i)/(1 - h_i) \quad (4)$$

Markup Rate = (Selling Price - Cost Price) / Cost Price * 100%

$$d_i = (b_i - e_{ij}) * e_{ij} * 100\% \quad (5)$$

Profit = Max (Sales - Cost)

$$l_i = \sum_{i=1}^n \max (b_i - c_i) \quad (6)$$

Sales = Quantity Purchased * Price

$$b_i = g_i * w_i \quad (7)$$

A line chart is created using regression analysis to illustrate the correlation between the total sales volume and the cost markup pricing for different vegetable categories. This chart offers a more lucid depiction of the correlation between the total sales volume and cost markup pricing for each vegetable category, as illustrated in Figure 1 below.

Regression coefficients: -0.0003445923178329567

Intercept: 8.557450453648398

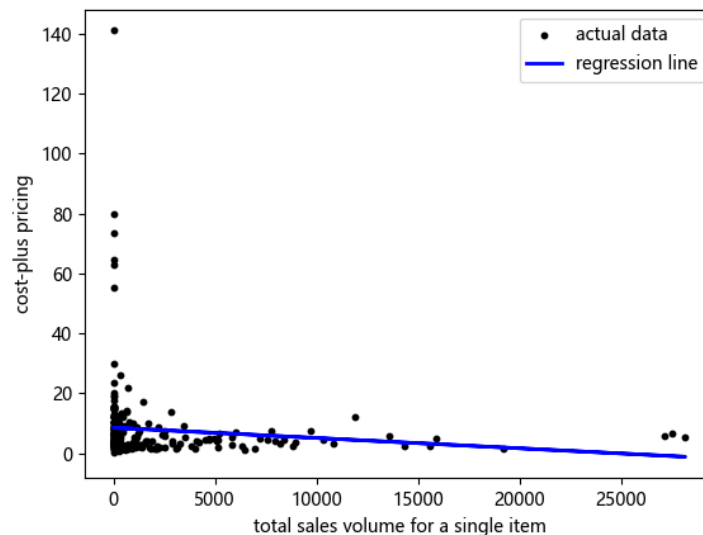


Figure 1. Relationship between total sales volume and cost-plus pricing for each individual vegetable

3.2. To optimize supermarket revenue, calculate the daily replenishment total for each vegetable category for the upcoming week (July 1-7, 2023) to maximize profits

During the model training process, the daily replenishment totals for each vegetable category from the initial 80% of the entire dataset were assigned to the training set, while the remaining 20% were allocated to the test set. The forecasting period ranged from July 1 to July 7, and the data was normalized. After several iterations, the optimal configuration was identified, with an autoregressive order p of 1, differencing order d of 2, moving average order q of 2, seasonal autoregressive order P of 1, seasonal differencing order D of 1, seasonal moving average order Q of 2, and a seasonal period s of 7. This configuration led to the minimum root mean square error (RMSE). The well-trained model was subsequently utilized to predict the replenishment quantities for each category of vegetables in the forthcoming week.

The anticipated replenishment quantities for each category of vegetables in the forthcoming week are outlined in Table 2 below.

Table 2. The daily replenishment quantities for various categories in the upcoming week (kg)

Time	Foliage Plants	Aquatic Rhizome Plants	Cauliflowers	Eggplants	Chili Peppers	Edible Fungi
2023/7/1	188.66	31.04	30.47	27.95	107.00	67.64
2023/7/2	179.48	28.41	28.41	27.41	102.97	65.32
2023/7/3	126.49	14.05	15.01	21.18	77.63	43.51
2023/7/4	119.99	13.84	14.21	18.57	75.96	40.83
2023/7/5	121.18	13.24	13.85	19.02	75.57	46.62
2023/7/6	117.52	12.58	14.76	18.96	77.35	44.97
2023/7/7	132.00	18.95	19.18	22.93	88.28	52.92

The visual representations of the projected data for various categories in the upcoming week are depicted in Figures 2 to 7 below.

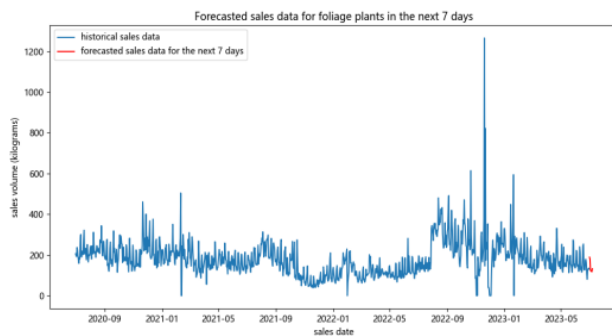


Figure 2. Foliage Plants

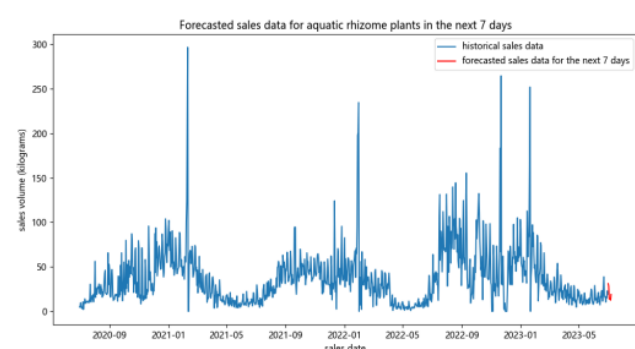


Figure 3. Aquatic Rhizome Plants



Figure 4. Cauliflowers

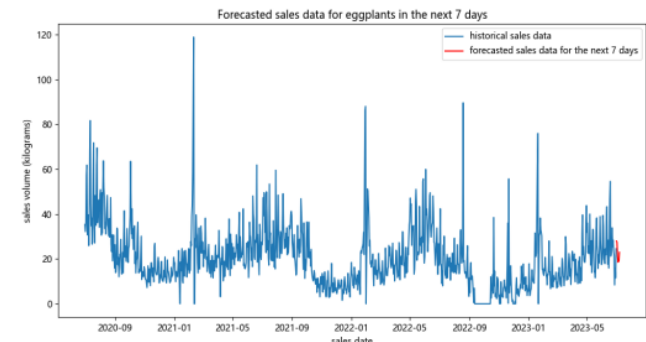


Figure 5. Eggplants

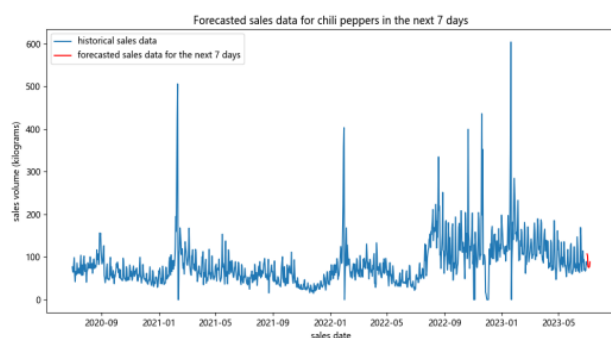


Figure 6. Chili Peppers

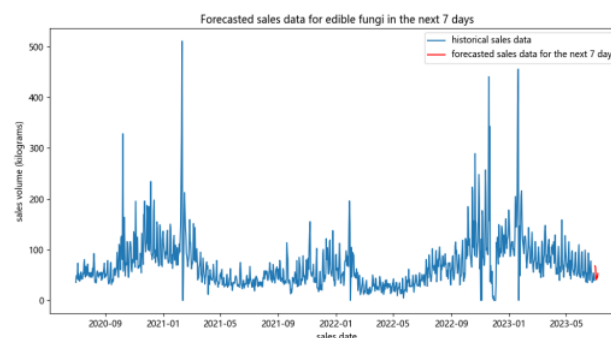


Figure 7. Edible Fungi

4. Conclusions

This paper examines the inventory replenishment and pricing challenges associated with the short shelf life of specific vegetable products in contemporary fresh convenience stores, particularly in the context of seasonal factors. The establishment of a linear relationship between total sales volume and cost markup pricing is achieved through the introduction of linear regression equations and their combination with the cost markup pricing formula. The collected data is used as the training set to construct a seasonal autoregressive moving average model for predicting the daily replenishment totals and pricing strategies for different vegetable categories in the upcoming week.

The graphical results reveal the following insights: ① There exists a significant correlation between the restocking amounts of different types of vegetables and the seasonal variations. Utilizing seasonal indices for forecasting can effectively minimize procurement expenses and improve overall revenue. ② A modest reduction in pricing has the potential to substantially boost the overall sales volume of various vegetable categories.

This study exclusively examines the influence of time on strategies for replenishment. Subsequent studies could investigate the impact of various factors, including regional variations, consumption patterns among diverse consumer demographics, and unexpected events, on strategies for replenishing convenience stores.

References

- [1] Zhao Ling, Liu Zhixue. Joint replenishment and pricing policy with customer returns and fixed costs [J]. *Operations Research and Management*, 2022, 31(06): 105-110+124.
- [2] Chen Jun, Kang Sha. Joint decision on pricing and inventory replenishment of agri-food with dual-channel sales [J]. *Industrial Engineering*, 2023, 26(03): 39-46.
- [3] Liu Yizheng, Ai Xueyi. Research on the joint replenishment and delivery problem of fresh product based on RFID technology and its investment decision-making model [J]. *Logistics Engineering and Management*, 2023, 45(07): 1-7+35.
- [4] Cui Ligang, Li Yali, Liu Jinxing, et al. Joint replenishment and pricing decision of multi-item considering the investment of fresh preservation technology [J]. *Industrial Engineering and Management*, 2023, 28(03): 17-26.
- [5] Liu Rui. Research on models and algorithms for multi-product joint replenishment and delivery scheduling problem [D]. Huazhong University of Science and Technology, 2017.
- [6] Ai Xueyi, Zhang Jinlong, Xu Haoxuan, et al. Optimal joint replenishment policy for multiple non-instantaneous deteriorating products with deterioration-time-dependent demand [J]. *Control and Decision*, 2018, 33(04): 671-678.
- [7] Wang Qiang. Research on inventory and pricing decisions considering dual replenishment models and reference price effects [D]. University of Science and Technology of China, 2020.

- [8] Ben Xiangming, Qin Beijin, Xiong Jianqiao, et al. Automatic replenishment intelligent system [J]. Electronic World, 2020, (14): 11-12.
- [9] Liu Yang. Research on demand forecast and replenishment of S chain drugstore [D]. Shandong University of Finance and Economics, 2021.
- [10] Le Yi. Research on models and algorithms for joint replenishment and delivery problems of multiple fresh products [D]. Wuhan University of Science and Technology, 2022.