Forecasting Daily Replenishment Volume and Pricing Strategies for Vegetables Based on Time Series Models

Ziliang Yin [#, *], Yuxin Sang [#], Hongsu Li

College of Software, Henan Normal University, Xinxiang, China, 453007

* Corresponding Author Email: yinzlrun@163.com

[#]These authors contributed equally.

Abstract. Against the backdrop of today's rapid socio-economic development, retail supermarkets are faced with increasingly complex sales management challenges, especially in the area of vegetable sales. As a seasonal and perishable commodity, the accurate forecasting of vegetable sales volume and the effective formulation of pricing strategies are crucial to the operational efficiency of retail supermarkets. The relationship between total daily sales and cost-plus pricing is analyzed by quantifying the total daily sales, average daily sales price per unit, the degree of daily sales discounts, and the daily sales shipping loss metrics for each category. The total sales and total pricing data of each vegetable category from July 1 to 7 in each of the past years were screened to forecast the total sales and pricing from July 1 to 7 in the year 2023 by using the ARIMA model in the time series analysis method. Based on the forecast results, the replenishment volume and pricing strategy for the day are adjusted to achieve the goal of maximizing revenue.

Keywords: Descriptive Statistical Analysis, Quantitative Relationship, ARIMA Model, Autocorrelation Analysis, Time Series Analysis.

1. Introduction

With the rapid development of the social economy and the continuous change in consumers' consumption habits, retail supermarkets are facing more and more challenges in their daily business activities, especially in the sales and management of vegetable commodities. As vegetables are an important part of daily consumer goods, fluctuations in their sales volume directly affect the operating efficiency of supermarkets. Therefore, how to scientifically and reasonably predict the sales volume of vegetables and formulate an effective pricing strategy has become the key to enhancing the revenue of superstores.

This study focuses on the ARIMA time series model, which is used to analyze and forecast trends, seasonality, and periodicity in time series data and can help remove these effects for accurate forecasting and anomaly detection. A time series ARIMA model-based approach was used to forecast the total sales and pricing of retail supermarkets from July 1 to 7, 2023.ARIMA model has been widely used in several fields. Hu used the ARIMA model in the stability prediction analysis of strong rainfall on slopes [1]; Zhai et al applied the ARIMA model in mine disaster prediction [2]; Yang et al [3], on the other hand, used the ARIMA model for urban economy and population prediction; Xie used ARIMA model in the economic analysis of the GDP of Shaanxi Province to analyze and predict the GDP of Shaanxi Province [4], and Yang proposed the petroleum price prediction model based on ARIMA-GM and empirical analysis [5].

In conjunction with the characteristics of the data on total vegetable sales by category, the relationship between total sales and cost-plus pricing was analyzed by first quantifying the total daily sales, the average daily unit price, the degree of daily sales discounts, and the daily sales shipping loss indicators for each category; Next, the total sales volume and total pricing of each vegetable category from July 1 to 7 in each year from 2020 to 2022 were screened, and the total daily replenishment volume and pricing strategy from July 1 to 7 in 2023 were predicted using the time series analysis (ARIMA) method and the daily revenue of the superstore for each of these 7 days was calculated based on the prediction results, which then provided feedback for adjusting the daily replenishment volume and the pricing strategy on these 7 days to achieve the maximum revenue.

2. Data Collection and Processing

2.1. Data preprocessing

Firstly, from the dataset (data source: http://www.mcm.edu.cn/), this paper filtered out the data related to the sales unit price and wholesale unit price of vegetables during July 1-7 in the three years from 2020-2022 and according to the category Divide it and calculate the total sales and total revenue of each vegetable unit within 7 days in each year as two new columns.

2.2. Quantitative indicator relationships

The study of the relationship between total sales and cost-plus pricing for each vegetable category begins with the quantification of the total sales of vegetables in the category, the average unit price, the degree of discount on the daily sales of the category, the transportation loss of vegetables, and the average wholesale unit price as specified below:

Let the total sales corresponding to day t be x_t, then

$$\mathrm{x}_t = \sum_{i=1}^{i=N} x_{i,t} - \sum_{i=1}^{i=N} y_{i,t} \tag{1}$$

The average unit price P_t for normal sales, where

$$P_t = \frac{\sum_{i=1}^{i=N} x'_{i,t} P_{i,t}}{\sum_{i=1}^{i=N} x'_{i,t}} \tag{2}$$

The average unit price of a discount sale is q'_t, where

$$q'_t = \frac{\sum_{i=1}^{i=N} x''_{i,t} q'_{i,t}}{\sum_{i=1}^{i=N} x''_{i,t}} \tag{3}$$

The degree of discount is k_t, where

$$k_t = \frac{q_t}{p_t} \times 100\% \tag{4}$$

The average wholesale price was u_t, of which

$$u_t = \frac{\sum_{i=1}^{i=N} (x_{i,t} - y_{i,t}) w_{t,i}}{x_t} \tag{5}$$

The quantification of the indicators in each of the six categories can be derived from the above formula:

Relationship between total revenue and total sales: There is a direct relationship between total revenue and total sales, with total revenue equal to total sales multiplied by the average selling price.

Profit about total revenue and costs: profit equals total revenue minus total costs, and maximizing profit is the goal of the superstore.

The relationship between selling price and cost-plus pricing: the selling price of an item is usually based on the cost of goods brought in plus a certain cost-plus rate.

Relationship of total cost to total sales and cost price: Total cost is equal to total sales multiplied by cost price.

It follows that there is a direct relationship between total cost and total sales.

3. Forecasting Vegetable Sales Trends

3.1. ARIMA modeling

3.1.1. Observation of time series smoothness

Since the ARIMA model requires the time series to be smooth, if the time series is not smooth, it is necessary to perform a differencing operation until the data becomes smooth. Smoothness can be judged by looking at whether the mean and variance of the time series change over time. Here the rolling statistics of vegetable sales over time series are plotted to demonstrate its smoothness. As shown in Figure 1:

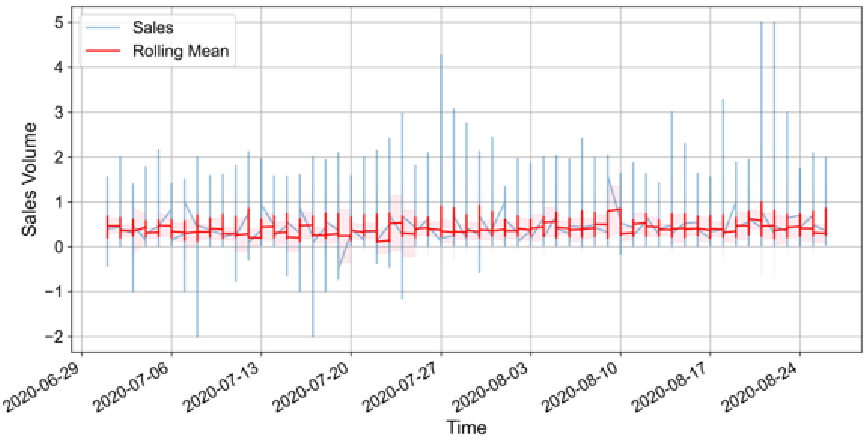

Figure 1: Rolling volume chart based on sales volume.

3.1.2. Determine the differential order d

The difference order d is determined by looking at the autocorrelation (ACF) and partial autocorrelation (PACF) plots of the time series. Here the time series-based ACF and PACF plots of vegetable sales and prices are plotted. As shown in Figure 2:

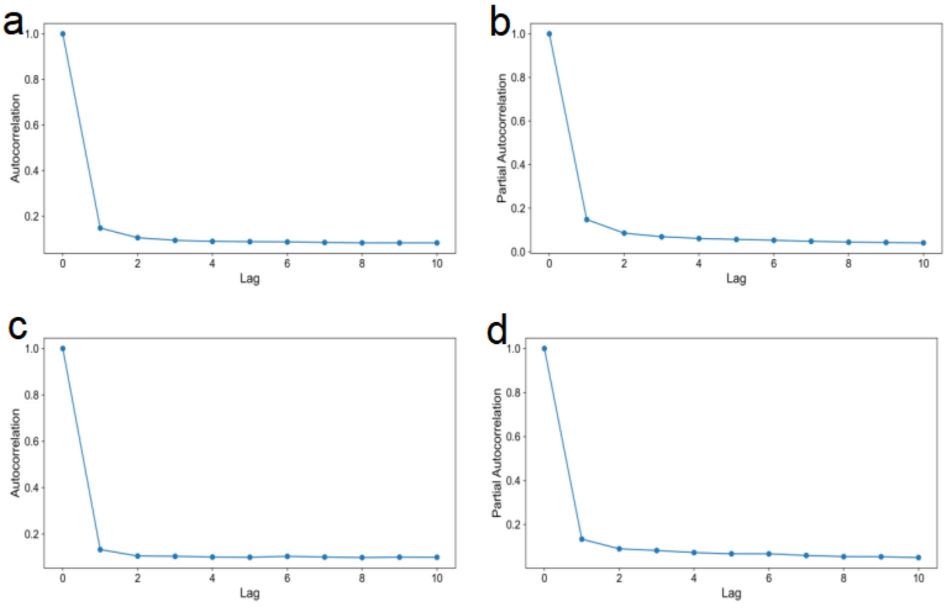

Figure2: ACF charts and PACF charts. a: Sales-based ACF Charts;
b: Sales-based PACF charts; c: Prices-based ACF charts; d: Prices-based PACF charts.

Observation of the above figure shows that the vegetable sales and price-based time series ACF show a strong positive correlation at the first few lagged orders and the PACF shows truncation at the same lagged orders, which results in the need for multiple differencing until the time series becomes smooth.

3.1.3. Determine the autoregressive order p and the moving average order q

Where the autoregressive order p is the first truncated lag order on the ACF plot and the moving average order q is the first truncated lag order on the PACF plot. There are significant truncated tails on both the ACF and PACF plots above, so consider trying some candidate models such as ARIMA(p, 0, q) or ARIMA(0, 0, q), in addition to trying to use an information criterion (AIC or BIC) to help select the appropriate model. After several experiments by trying different values of (p, d, q), the model with the lowest AIC (or BIC) value is selected. At this point, the optimal model parameters are ARIMA (1, 1, and 3). In the following, the specific process of modeling is shown.
ARIMA(p,d,q) modeling:

For a non-smooth sequence, if a smooth time series can be obtained after a few differences, such a sequence is said to be a single integer sequence. Let y_t be a single integer series of order d, denoted as: $y_t \sim I(d)$, then

$$w_t = \Delta^d y_t = (1 - l)^d y_i \tag{7}$$

Is a smooth series, then $w_t \sim I(0)$, and so an ARIMA (p, q) model can be built for w_t:

$$w_t = c + \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-p} \tag{8}$$

If the time series { y_t } is an ARIMA(p, q) process after d times of differencing, the original time series is said to be a p-order autoregressive, d-order rounding, q-order moving average process, denoted as ARIMA(p, d, q), with d representing the number of differencing.

3.2. Fitting of the model

The ARIMA model was fitted based on the determined values of (p, d, q), and the model fit and the nature of the residuals were checked to see if the model assumptions were met. The sales and revenues of the six types of vegetables from July 1 to 7, 2020-2022 were analyzed and visualized based on the ARIMA time series model. As shown in Figure 3 :(Take aquatic rhizomes as an example)

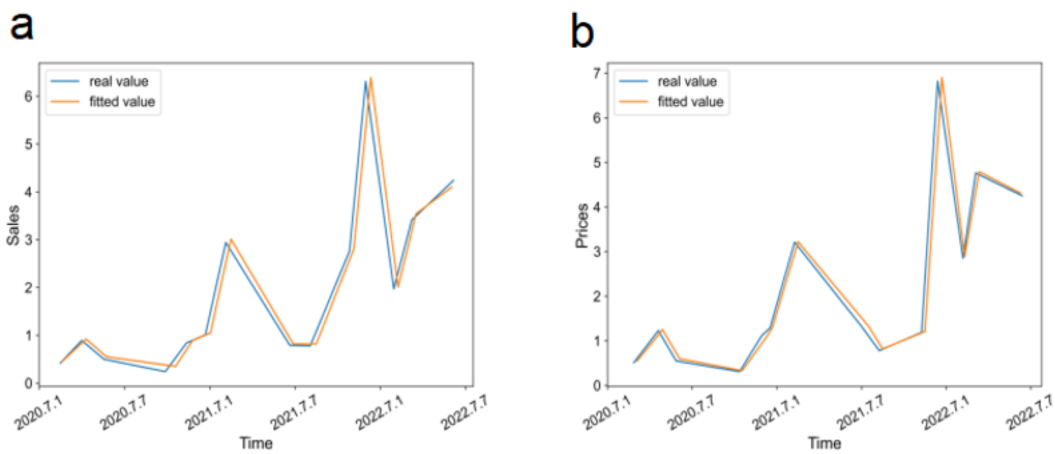

Figure 3: ARIMA model fitting plot. A: Fitted chart based on sales; b: Fitted chart based on prices.

Observing the above figure, fitting the data on the AMIRA model with defined parameters, the difference between the true and fitted values is small, and the residual series shows stochasticity, smoothness, and absence of autocorrelation, which are key indicators that highlight that the ARIMA model is well fitted. This further illustrates the strong correlation between relationship between vegetable sales and cost-plus pricing for each category, the feasibility of forecasting future vegetables in the first week of July each year through the sales volume and price of vegetables from July 1 to 7 from 2020 to 2022, and the guidance for calculating the total daily replenishment volume and pricing strategy for each vegetable category under the premise of guaranteeing the maximum revenue for the superstores.

Based on the above analysis, the projected replenishment volumes and revenues for each type of vegetable from July 1 to 7, 2023 are shown in Table 1 and Table 2:

Table 1: Aquatic Roots, Edible Mushrooms, and Capsicum Recharge and Revenue, July 1-7, 2023.

Time	Aquatic rhizomes		Edible mushroom		Capsicum	
	Replenishment (kg)	Income (yuan)	Replenishment (kg)	Income (yuan)	Replenishment (kg)	Income (yuan)
2023/7/1	44.98	475.43	43.26	196.36	38.03	507.06
2023/7/2	46.89	494.72	43.26	169.97	36.19	520.26
2023/7/3	48.8	514.02	43.26	141.95	34.35	529.45
2023/7/4	50.71	533.31	43.26	141.6	32.51	535.84
2023/7/5	52.62	552.6	43.26	122.61	30.66	540.29
2023/7/6	54.53	571.89	43.26	103.6	28.82	543.38
2023/7/7	56.44	591.19	43.26	93.43	26.98	545.53

Table 2: Recharge and Revenue of Cauliflower, Eggplant and Philodendron, July 1-7, 2023.

Time	Cauliflower		Eggplant		Philodendron	
	Replenishment (kg)	Income (yuan)	Replenishment (kg)	Income (yuan)	Replenishment (kg)	Income (yuan)
2023/7/1	43.92	398.45	32.38	169.59	130.48	683.72
2023/7/2	43.92	389.21	33.53	185.3	126.91	644.67
2023/7/3	43.92	379.97	33.53	185.3	123.34	605.62
2023/7/4	43.92	370.73	33.53	185.3	119.77	566.56
2023/7/5	43.92	361.49	33.53	185.3	116.2	527.51
2023/7/6	43.92	352.25	33.53	185.3	112.63	488.46
2023/7/7	43.92	343.00	33.53	185.3	109.07	449.41

Based on the data presented in Tables 1 and 2, it can be concluded that the pricing strategies for each type of vegetable when the supermarket returns are maximum from July 1 to 7, 2023 are shown in Table 3 below:

Table 3: Pricing Strategies for Various Categories of Vegetables When Supermarket Revenue is Maximized, July 1-7, 2023.

Time	Aquatic rhizomes.	Edible mushroom	Capsicum	Cauliflower	Eggplant	Philodendron
	Pricing (yuan/kg)	Pricing (yuan/kg)	Pricing (yuan/kg)	Pricing (yuan/kg)	Pricing (yuan/kg)	Pricing (yuan/kg)
2023/7/1	10.57	4.54	13.33	9.07	28.64	5.24
2023/7/2	10.55	3.93	14.38	8.86	5.53	5.08
2023/7/3	10.53	3.28	15.41	8.65	5.53	4.91
2023/7/4	10.52	3.27	16.48	8.44	5.53	4.73
2023/7/5	10.5	2.85	17.62	8.23	5.53	4.54
2023/7/6	10.49	2.39	18.85	8.02	5.53	4.34
2023/7/7	10.47	2.16	20.22	7.81	5.53	4.12

4. Conclusions

This study validates the effectiveness of quantitative analysis and time series forecasting models in the formulation of revenue maximization strategies for superstores by thoroughly analyzing the relationship between total vegetable sales and cost-plus pricing, and successfully forecasting the total daily replenishment and pricing strategies for each vegetable category from July 1 to 7, 2023, using the ARIMA time series model.

The relationship between total daily sales and cost-plus pricing for each vegetable category was investigated by quantifying the total daily sales, average daily sales unit price, degree of daily sales discount, and daily sales shipping loss metrics for each category. The total daily replenishment and pricing strategies for each vegetable category from July 1 to 7, 2023 were successfully predicted based on an ARIMA time series model. During the model fitting process, it was observed that the fitted values were in good proximity to the true values, and the residual series exhibited randomness, smoothness, and no autocorrelation, which verified the validity and reliability of the adopted model in sales forecasting.

The results of this paper validate the effectiveness of quantitative analysis and time series forecasting models in the development of revenue maximization strategies for superstores. By analyzing sales data in depth and applying the ARIMA model for forecasting, we provide a viable sales management method for superstores. This research idea and framework demonstrate the feasibility of applying quantitative analysis and time series forecasting models in the field of market analysis, which provides an important reference for the business decisions of superstores.

References

[1] Xin Hu. Stability prediction analysis of strong rainfall on slopes based on time series ARIMA model [J]. Pearl River Water Transport, 2024, (04):74-76.

[2] Xiaowei Zhai, Xiang Zhou, Bobo Song, et al. Research on mine disaster prediction method based on ARIMA model [J/OL]. Coal Technology, 1-6[2024-05-31].

[3] Yang Yang, Dingsheng Tian, Bao'an Zhang, et al. Research on urban economy and population forecasting based on ARIMA model [J]. Comprehensive Transport, 2023, 45(11):79-85+97.

[4] Nini Xie. Application of time series model in economic analysis - GDP analysis and forecasting in Shaanxi Province [J]. Sino-Arab Science and Technology Forum (in English and Chinese), 2024, (02):42-46.

[5] Lanqing Yang. Oil price forecasting model based on ARIMA-GM and empirical analysis [J]. Modern Industrial Economy and Information Technology, 2023, 13(11):28-31.

[6] Haoxu Yang. Research on the correlation analysis of vegetable price and sales volume--Taking the example of oleaginous vegetables [J]. Food Safety Guide, 2018, (21):179-181.

[7] Liyuan Ning. Pricing strategies and techniques of green vegetables [J]. Guangdong Sericulture, 2020, 54(07):111-112.

[8] Yanhui Zhou, Zuomiao Shi, Guohua Zhang, et al. pricing strategy of vegetable seeds [J]. Yangtze River Vegetable, 2013, (01):56-57.

[9] Junbo Zhuang, Xiaohong Cai, Zhen Shi, et al. Weekly cigarette sales prediction model based on fitting market state characteristics[J]. Mall Modernisation, 2024, (01):1-4.

[10] Zelin Lou, Junhong Zheng, Lili He. Research on sales volume prediction based on improved SSA-LSTM [J]. Computer Age, 2023, (10):50-53+58.