Information Systems and Economics (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/infse.2024.050110
ISSN 2523-6407 Vol. 5 Num. 1

Research on Vegetable Replenishment Strategy Based on Time Series Analysis

Tongao Zhang[1,†,*], Zhongkai Zhang[1,†], Yunsheng Chi[1,†]

[1]College of New Energy, China University of Petroleum, Qingdao, China
[†]These authors also contributed equally to this work
[*]Corresponding author: 2930196638@qq.com

Keywords: Vegetable restocking, Pearson's correlation coefficient, time series analysis, Markov forecasting

Abstract: This paper centers on the problem of how superstores can accurately determine the replenishment number of several types of vegetables to ensure their freshness, and utilizes the Pearson correlation coefficient method, the time series analysis method, and the Markov prediction method to conduct the research. Firstly, the distribution pattern of sales volume of vegetable categories and their interrelationships were studied by preprocessing vegetable sales data, and correlation coefficients were calculated to assess the correlation between these categories; secondly, the relationship between the total sales volume of vegetable categories and the cost-plus pricing was analyzed, and the daily replenishment volume and pricing strategy for the coming week were predicted by using the time-series method; lastly, the combination of several factors was used to analyze the specific time period of the high total sales price individual items at a given time was analyzed, and Markov forecasting was applied to determine the optimal replenishment volume and pricing strategy. The model prediction results of this study show the practicality and significance for ensuring the freshness of vegetables and maximizing the benefits of supermarkets.

1. Introduction

As an indispensable food in people's daily life, the freshness and quality of vegetables have always been the focus of consumers' attention. With the development of social economy, people's demand for vegetables is increasing, expecting to buy fresher products. In fresh food supermarkets, the freshness period of vegetables is generally shorter, and their appearance and quality tend to gradually decline with time. Therefore, it is particularly important to analyze the past sales data to grasp the sales distribution pattern of each category of vegetables and determine the future replenishment volume and pricing strategy accordingly. This study aims to solve the pricing strategy and replenishment problem of vegetable items, and to explore in depth the sales correlation among vegetable categories, the relationship between total sales volume and cost-plus pricing through layer-by-layer model solving. This paper will use time series analysis algorithms to predict the total replenishment volume and pricing strategy for the next seven days based on data from April to June 2023, and based on this, specific individual items will be selected for July 1 replenishment and pricing to maximize benefits. Finally, the paper will explore other potential factors that influence vegetable replenishment and pricing and explain the reasons for their selection.

2. Relevance Analysis

2.1 Data processing

In the data preprocessing stage of this paper, facing the huge and partially duplicated dataset, this paper adopts the methods of aggregation processing and matching processing to simplify the data structure. Using Excel, this paper summarizes the vegetable sales data from July 1, 2020, to June 30, 2023, and divides the daily sales volume by category and individual product. At the same time, to solve the problem of missing individual product codes with corresponding names and categories in the data, this paper utilizes Excel's VLOOKUP function. This step not only enhances the integration and readability of the data, but also provides a clear and effective database for in-depth analysis [1]. The relationship between the sales volume of each vegetable category and time was plotted, the final distribution pattern of sales volume for each vegetable category was obtained as shown in Fig. 1.

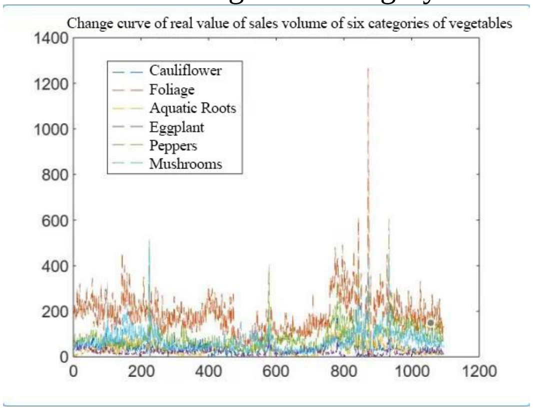


Figure 1: Final distribution pattern of sales volume by vegetable category

The three-year sales totals for each vegetable category were processed to obtain a graph of the total sales share of each category as shown in Fig. 2 below:

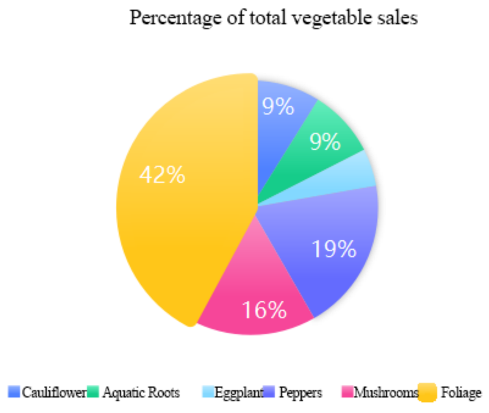


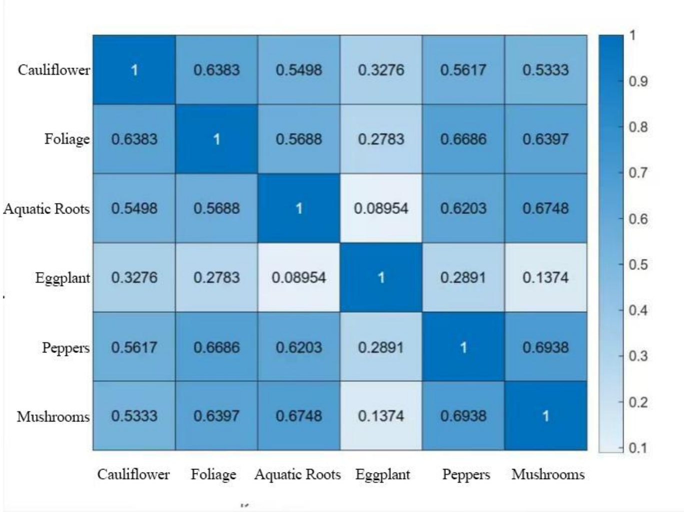
Figure 2: Total Sales Share by Category

By analyzing the distribution pattern of sales volume of each category of vegetables, it is understood that the sales volume of vegetable categories shows a seasonal cycle pattern with time. Moreover, during these three years, the total sales volume of leafy and flowering vegetables was the largest and the sales volume of eggplant vegetables was the smallest.

2.2 Correlation analysis

In this paper, the Pearson correlation coefficient is used to measure the linear correlation between six different vegetable categories. This method is based on the covariance matrix of the data to assess the strength of the relationship between two variables. Specifically for this study, these six categories include foliar, cauliflower, eggplant, pepper, aquatic root, and edible mushrooms. This paper calculates the Pearson correlation coefficients between these categories by comparing their total sales volume two by two to determine the strength of the correlation. To visualize these correlations more, this paper also draws a heat map. In addition, based on the values of the correlation coefficients, this paper classifies the degree of correlation into extremely strong correlation, strong correlation, moderate correlation, weak correlation, and extremely weak correlation or no correlation, thus providing a clear criterion for judging the analysis in this paper. This method effectively reveals the interrelationships among different vegetable categories and the strength of their correlations [2].

Their Pearson's correlation coefficients were calculated two by two and heatmaps were plotted and the results were obtained as shown in Fig. 3:



	Cauliflower	Foliage	Aquatic Roots	Eggplant	Peppers	Mushrooms
Cauliflower	1	0.6383	0.5498	0.3276	0.5617	0.5333
Foliage	0.6383	1	0.5688	0.2783	0.6686	0.6397
Aquatic Roots	0.5498	0.5688	1	0.08954	0.6203	0.6748
Eggplant	0.3276	0.2783	0.08954	1	0.2891	0.1374
Peppers	0.5617	0.6686	0.6203	0.2891	1	0.6938
Mushrooms	0.5333	0.6397	0.6748	0.1374	0.6938	1

Figure 3: Heat map of Pearson's correlation coefficient

According to the above figure, this paper concludes that cauliflower, foliage, aquatic roots, chili peppers, edible mushrooms can be viewed as a category with more than moderate correlation between the two of them, and eggplant is viewed as a separate category, which has a weak correlation with the rest of the vegetable categories.

3. Category pricing and replenishment strategies

3.1 Modeling

Scatter plots of daily sales volume versus unit price for each category of vegetables were made and fitted to finally obtain the relationship between sales volume and pricing in the case of edibles and chili peppers as shown in Fig. 4:

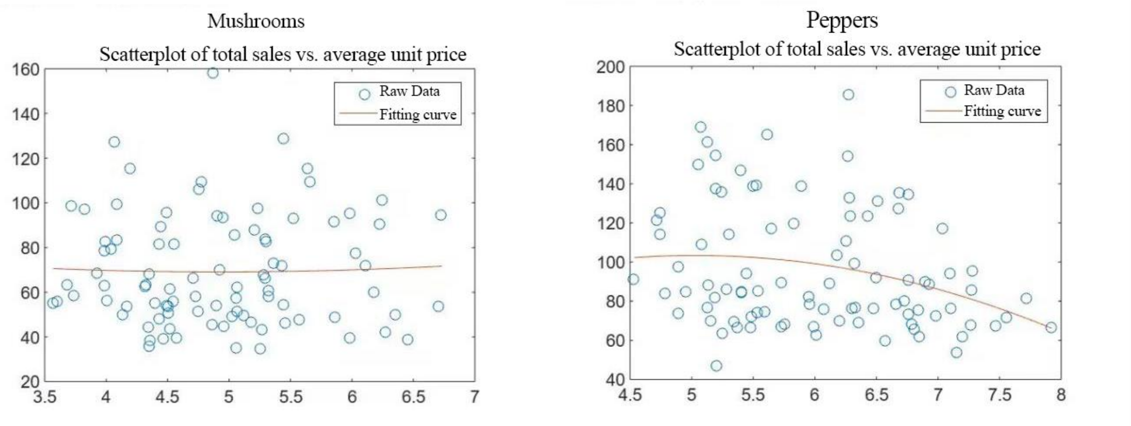


Figure 4: Scatterplot between total sales and pricing for the vegetable category and the corresponding fitted curve equation

From the above analysis, the relationship between total sales volume and unit price for each vegetable category over a three-month period is obtained as follows:

Relationship between total quantity of edible mushrooms sold and unit price:

$$\sigma_{1k} = 0.79432 * w_{1k}{}^\wedge 2 - 7.833 * w_{1k} + 88.405 \tag{1}$$

Relationship between total sales volume and unit price of chili peppers:

$$\sigma_{2k} = 0.79432 * w_{2k}{}^\wedge 2 - 7.833 * w_{2k} + 88.405 \tag{2}$$

Relationship between total eggplant sales and unit price:

$$\sigma_{3k} = 26.134 * w_{3k} - 1.8612 * w_{3k}{}^\wedge 2 - 69.902 \tag{3}$$

Relationship between total aquatic rootstock sales and unit price:

$$\sigma_{4k} = 0.82627 * w_{4k} - 0.055135 * w_{4k}{}^\wedge 2 + 14.625 \tag{4}$$

Relationship between total sales and unit price of foliage categories:

$$\sigma_{5k} = 26.943 * w_{5k} - 2.5083 * w_{5k}{}^\wedge 2 + 95.496 \tag{5}$$

Relationship between total cauliflower sales and unit price:

$$\sigma_{6k} = 0.057302 * w_{6k}{}^\wedge 2 - 3.4352 * w_{6k} + 54.313 \tag{6}$$

In performing time series analysis, the Autoregressive Moving Average (ARIMA) model has been chosen in this paper. Time series analysis, as a statistical technique, focuses on time-ordered data series and aims to reveal patterns in the data over time and to be used for forecasting. The ARIMA model assumes that the current values of a time series can be viewed as a linear combination of its past values and disturbing values [3].

The model divides the time series into an autoregressive part and a moving average part. The autoregressive part considers the past values of the series, while the moving average part deals with the past disturbances. In the model, the orders p and q represent the autoregressive and moving average orders, respectively, while the coefficients φ and θ are estimates of the autoregressive and moving average coefficients, respectively. In addition, residuals or white noise series in the model are those random fluctuations that cannot be explained by the model [4].

The following is an example of the pattern of change in the total daily sales of cauliflower vegetables over time to be analyzed as in Fig. 5 and 6. Considering that since the distribution law of data over time may have an unsteady nature, this paper performs backward differencing on the data to attenuate its unsteady trend. A more reliable distribution law of the data over time is obtained.

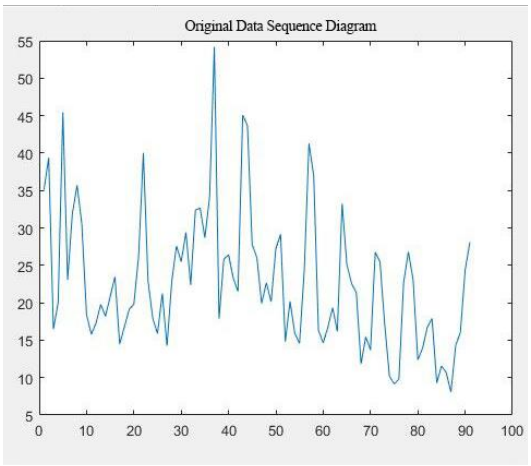


Figure 5: Pre-differential sales volume sequence diagram

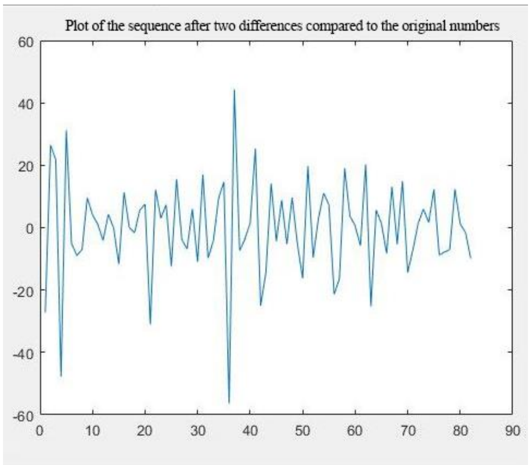
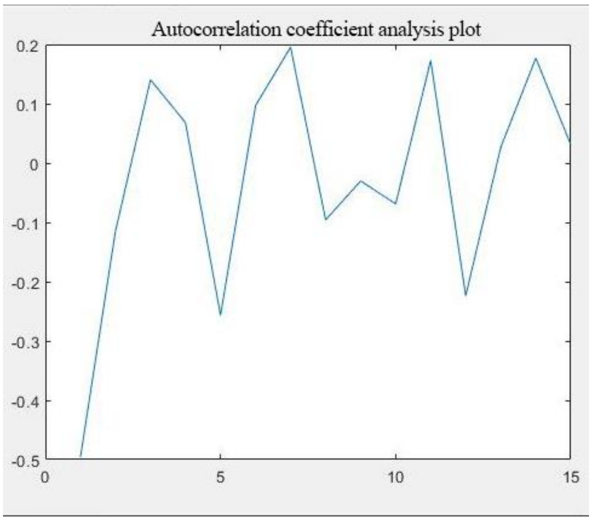


Figure 6: Sequence diagram of sales volume after differencing

The autocorrelation coefficient is calculated based on the autocovariance and variance of the total daily sales of a specific vegetable category, reflecting the degree of correlation at different time points in the time series data. The partial correlation coefficient, on the other hand, is estimated by modifying the ARIMA model to help understand the direct relationship between time points in the time series data and eliminate the influence of other factors. In this paper, the optimal order of the autoregressive moving series model, i.e., the combination of p and q values that minimize the AIC value in the model, is determined using the Akaike informativeness criterion (AIC). In addition, this study proposes a model for predicting the total daily replenishment of each vegetable category in the coming week, the model is based on the relationship between the wastage rate and the amount of incoming goods, combined with the time series prediction model, to provide a scientific basis for the replenishment decision of vegetable goods as shown in Fig. 7 and 8. Through these methods, this paper aims to improve the accuracy of forecasting vegetable sales trends and provide solid data support for related decisions [5].



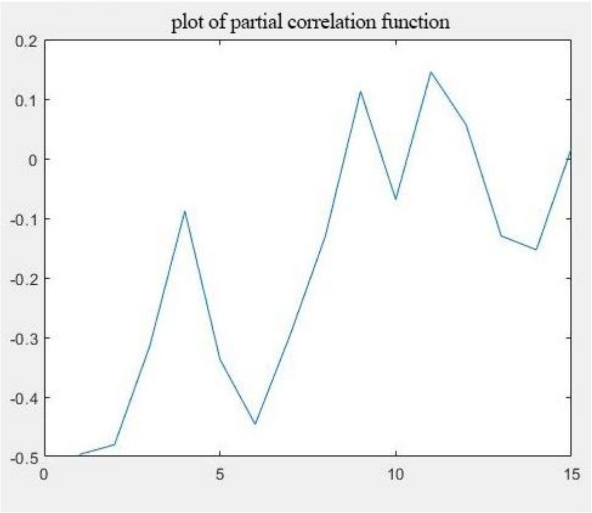


Figure 7: Autocorrelation coefficient plot Figure 8: Plot of partial correlation coefficients

3.2 Solving the model.

The total daily replenishment of each vegetable category in the first three months was used as a training set to build the prediction model. The data from 23 to June 30, 2023, is used as a test set to evaluate the reasonableness of the prediction model. The results obtained are shown in Fig. 9:

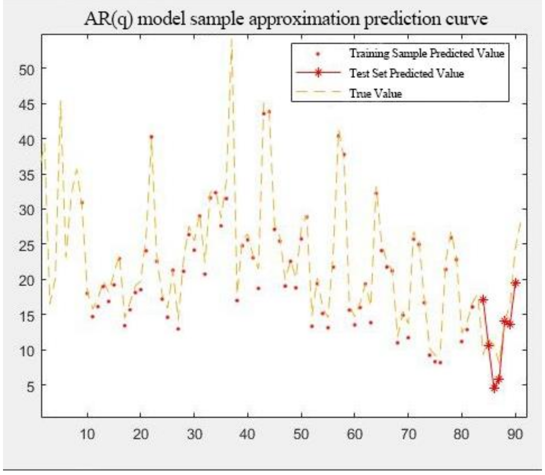


Figure 9: Test set test results for daily replenishment of cauliflower category

From the above figure, the model predicts more accurate predictions compared to the actual values. Therefore, this model is used to predict the total daily replenishment of each vegetable category in the coming week. After the above prediction model was evaluated and passed, it was used to predict the total daily replenishment and profit of cauliflower vegetable category in the coming week, and the results are shown in Fig. 10 and 11:

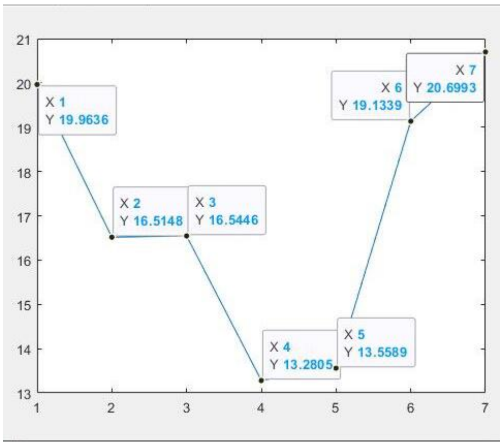


Figure 10: Forecast of daily replenishment in cauliflower category for the coming week

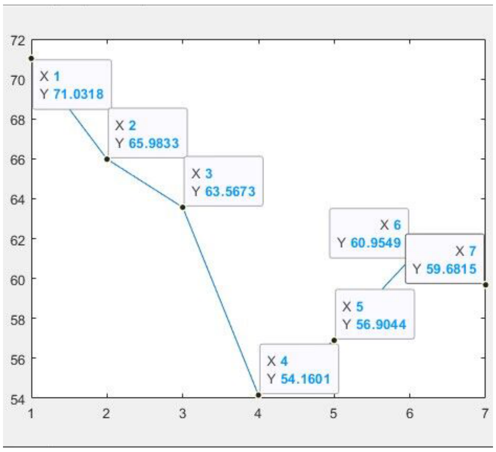


Figure 11: Forecast of daily profits for the cauliflower category for the week ahead

The predicted results of daily profit for the coming week for each vegetable category are given below in tabular form as Table 1:

Table 1: Predicted values of daily profits for each vegetable category in the coming week

	Cauliflower	Philodendron	Aquatic rhizomes	Eggplant	Capsicum	Edible mushroom
July 1	71.0318	207.384	46.1356	79.1513	191.088	77.718
July 2	65.9833	191.098	30.6889	57.8843	182.279	110.098
July 3	63.5673	171.719	43.298	51.1134	233.222	89.9729
July 4	54.1601	153.409	22.6873	60.3341	159.01	45.0576
July 5	56.9044	147.781	45.633	59.8486	187.887	57.4488
July 6	60.9544	165.796	37.1571	63.9477	170.773	79.9737
July 7	59.6815	172.656	37.3474	65.4893	173.44	68.9856

4. Individual product pricing and replenishment strategies

In this paper, the concept of Markov chain is introduced in the analysis and applied to predict replenishment, cost, and profit. A Markov chain is a stochastic sequence characterized by the fact that the future state of a system depends only on the current state and is independent of earlier states. This property is reflected in the fact that, given the current state, the probability distribution of the future state is independent of the past state [6].

In predicting the replenishment volume, cost and profit on July 1, this paper uses a Markov chain model to process the data of seven individual items. By analyzing the previous week's replenishment volume, cost, and profit, the paper calculates the probability that these individual items are in different states. This probability information is summarized in a table showing the probability distribution of each individual item in three different states (e.g., replenishment volume, cost, and profit). This approach not only provides predictions of future conditions, but also provides data support for developing effective replenishment and pricing strategies. By applying Markov chains and transfer probability matrices, this paper aims to improve the accuracy of forecasting and optimize resource allocation.

Using the meaning of the three states of each of the three metrics, replenishment, cost, and profit, as a basis, and multiplying each by their respective weights (derived probabilities), the final replenishment quantity is obtained, and the pricing strategy is shown in Table 2:

Table 2: Projected values of replenishment volume, cost, and profit for seven individual products

	Replenishment	costs	Margins	Pricing strategy
Amaranth greens	9.11993061	2.44370519	14.37160528	4.574993165
Brussels sprouts	13.59017277	2.43769731	20.66173881	4.342406266
Shanghai Youth	3.94650429	4.33318345	16.15560686	9.157554824
Snow fungus	6.06437934	3.37642856	14.05012264	5.971366775
Baby Chinese cabbage	10.65873	4.97452464	20.1131301	6.988040189
Sweet potato tip	4.60603146	3.36891871	10.42016757	5.940949592
Cabbage	6.56446359	2.27848849	15.64635328	5.085685273

5. Conclusions

The model constructed in this paper demonstrated efficiency in data preprocessing, ensuring data quality and simplifying subsequent operations. Through the application of Pearson's correlation coefficient, the model accurately assessed the relationship between the sales volume of different vegetable categories. Meanwhile, the adopted time series forecasting method has wide applicability and high accuracy, which makes the model more realistic and dependable in terms of cost, pricing, and sales volume. Nevertheless, the model has potential errors in the representativeness of the relationship between sales volume and pricing for certain vegetable categories and the stability of the time series forecasts. Looking ahead, the model has good potential for replication and can be adapted to different commodity categories to optimize replenishment and pricing strategies by adjusting sales volume data to bring maximum economic benefits to superstores.

References

[1] Huang Jia. Elementary usage of VLOOKUP function[J]. Computer Age, 2022, (10):121-122+126. DOI: 10.16644/j. cnki. cn33-1094/tp.2022.10.029

[2] Yuan-Shang Zhao,Wei-Fang Lin. Research on typical scenarios based on Pearson correlation coefficient fusion of peak density and entropy weight method [J]. China Electric Power, 2023, 56(05):193-202.

[3] Wang Ting, He Xiangfan. Prediction of bacillary dysentery incidence trend in Xinjiang based on seasonal autoregressive moving average model [J]. Public Health and Preventive Medicine, 2023,34(05):30-34.

[4] Dai Luping, Shen Jiayi, Zhang Feifei. An automatic energy and power demand forecasting model based on time series algorithm[J]. Automation Technology and Application, 2024,43(01):49-51+65.DOI:10.20033/j.1003-7241. (2024) 01-0049-04

[5] Shi M , Liao X , Yang H ,et al. Research on Vegetable Greenhouse Strategy Based on Multi-objective Distributed Constraint Optimization[C]//IEEE International Conference on Artificial Intelligence and Computer Applications. IEEE, 2021.DOI:10.1109/ICAICA52286.2021.9498123.

[6] Hou Jie, Wang Zhengyi. Research on the prediction of building electrical density in ecological parks based on gray Markov chain[J]. Modern Building Electricity, 2023, 14(11):6-10+59. DOI: 10. 16618/j.cnki.1674-8417.2023.11.002