Study on the pricing and replenishment of vegetable commodities based on the optimization model

Yongqiang Yin, Xuening Kong*, Runqi Cao

Inner Mongolia University Of Science & Technology, Inner Mongolia, China, 014010

* Corresponding Author Email: 18931936966@163.com

Abstract. Based on the problems of short shelf life, inventory control and pricing management, marketing, economics, operation research, and the control optimal theory, this paper establishes the regression prediction analysis model; establish the optimization model, formulate the total amount of daily replenishment; analyze the deficiencies of the model. Finally, the combination of the research results gives the replenishment and pricing decision to improve the revenue of the supermarket. In this study, the correlation degree between the two variables of each vegetable category and the single product sales volume was measured, and the time series analysis was used to identify the sales trend. Then according to the correlation analysis of descriptive statistics, using cosine similarity formula, analyze the data units, the total sales dimension normalization, using the method of standardize the correlation between item and category, using pearson correlation coefficient and Spearman correlation coefficient to study the relationship between category and sales, establish Spearman correlation coefficient to determine the final correlation coefficient, cluster analysis of each category, analysis of the correlation. By processing the relevant data, establish the ARIMA time series model, forecast on July 1,2023-July 7, the relevant data, using the adftest function time series stability test, fit the ARIMA time series model, using differential sequence to time series analysis, calculate vegetable sales and sales price of elastic demand function, finally find the optimal solution through the particle swarm algorithm.

Key words: Pearson Person Correlation, Cluster Analysis, ARIMA Time Series Model, Demand Price Elasticity.

1. Introduction

Fresh food refers to the primary products that can be placed in the store shelves without the initial processing process, such as necessary preservation and simple processing, such as cooking and production. Fresh vegetables have high water content, short preservation period, and easy to spoil, so fresh vegetables have timeliness requirements[错误!未找到引用源。]; According to the food itself, the safety requirement of consumers is varied, and the vegetable market is seasonal and seasonal. Therefore, how to make a decision on the basis of reducing product loss and cost control, the pricing and replenishment of each vegetable category on the day has become a major difficulty for fresh suppliers.

From the status of domestic research in recent years, the research on perishable products is few and scattered. From the current series of studies, most of the research technologies have been formed at home and abroad. Specifically for inventory management model and common, but almost all is based on the study of ordinary items, less specialized only for metamorphic research, metamorphic external characteristics for strong seasonal, short life cycle, the current domestic market metamorphic consumption scale, People's Daily life and the relationship between fresh goods inseparable. This study focuses on the inventory and sales of fresh food. On the basis of the existing research contents and results of ordinary goods and metamorphic products, we consider the uniqueness of fresh food, and make scientific and reasonable pricing and replenishment decisions[2].

From the demand side, there are many varieties of vegetables in the supermarket, and the supply varieties are abundant from April to October, but consumers have certain collocation in the selection of goods, so it is necessary to construct a reasonable sales combination to remove the restriction of sales space[3][3]. The purchase and trading time of vegetables is usually from 3:00 to 4:00 in the

morning. If the merchant does not clearly know the specific single product and the purchase price, the merchant needs to make the replenishment decision of each vegetable category on that day[4][4]. The pricing of vegetables generally adopts the "cost plus pricing" method, and the goods with transport loss and product phase difference are usually sold at a discount to reduce the loss rate. Reliable market demand analysis is crucial to replenishment decisions and pricing decisions[5][5][].

2. Exploration of the relationship and law of vegetable categories and single products

Due to different categories or different item may have certain correlation between link, so on the research data square experiment, mathematical correlation through mathematical model, focus on the establishment of the latter model, need to category and item, category and sales, item and sales of correlation analysis respectively[6][6]. Statistical correlation analysis of the number of vegetable categories and single products by MATLAB. The code operation results are shown below for the correlation analysis of the sales of main dishes in July. The correlation calculation method here is the cosine similarity formula as shown in Fig 1

$$\cos(\theta)=\frac{A\cdot B}{\|A\|\|B\|}=\frac{\sum_{i=1}^{n}A_i\times B_i}{\sqrt{\sum_{i=1}^{n}(A_i^2)}\times\sqrt{\sum_{i=1}^{n}(B_i^2)}} \tag{1}$$

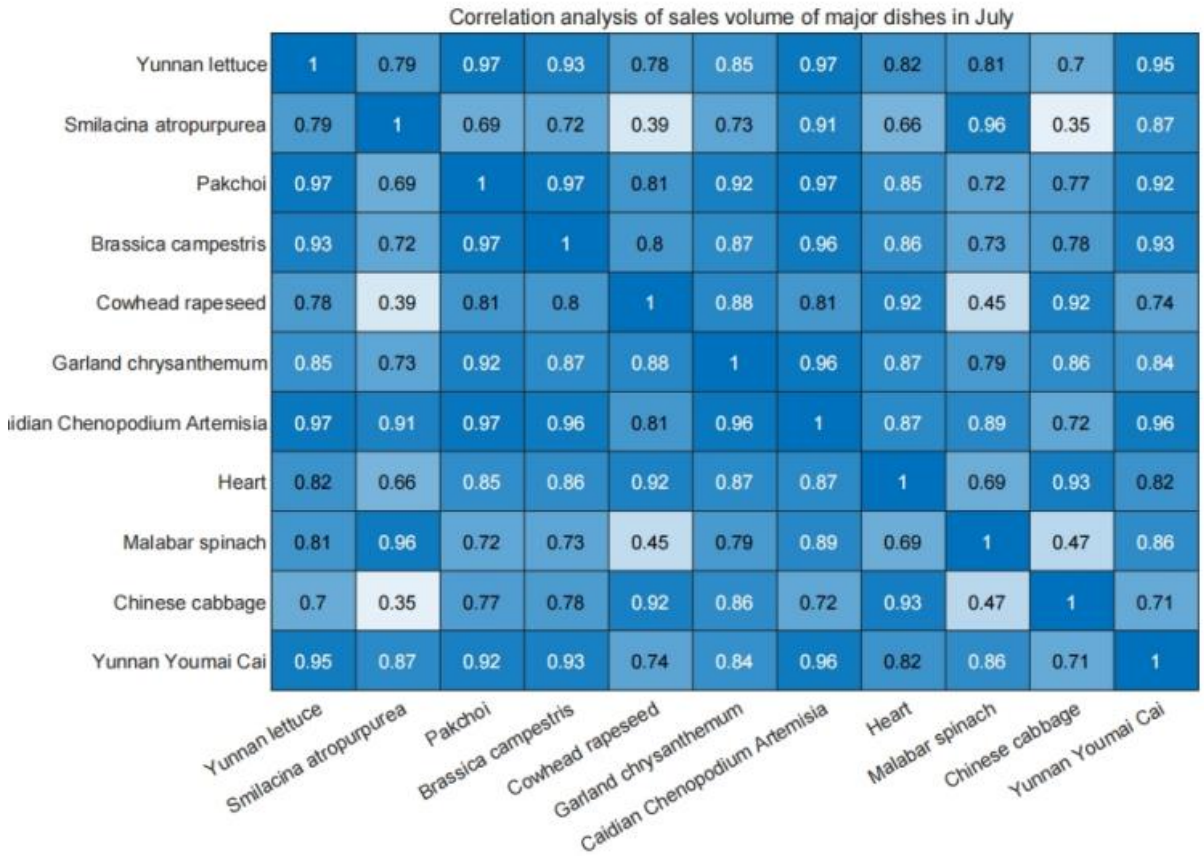

Fig. 1 Correlation analysis of the sales volume of main dishes in July

The jupter with Ancona and do the correlation analysis between category and category in jupter to obtain the following correlation heat map,as shown in Fig 2.

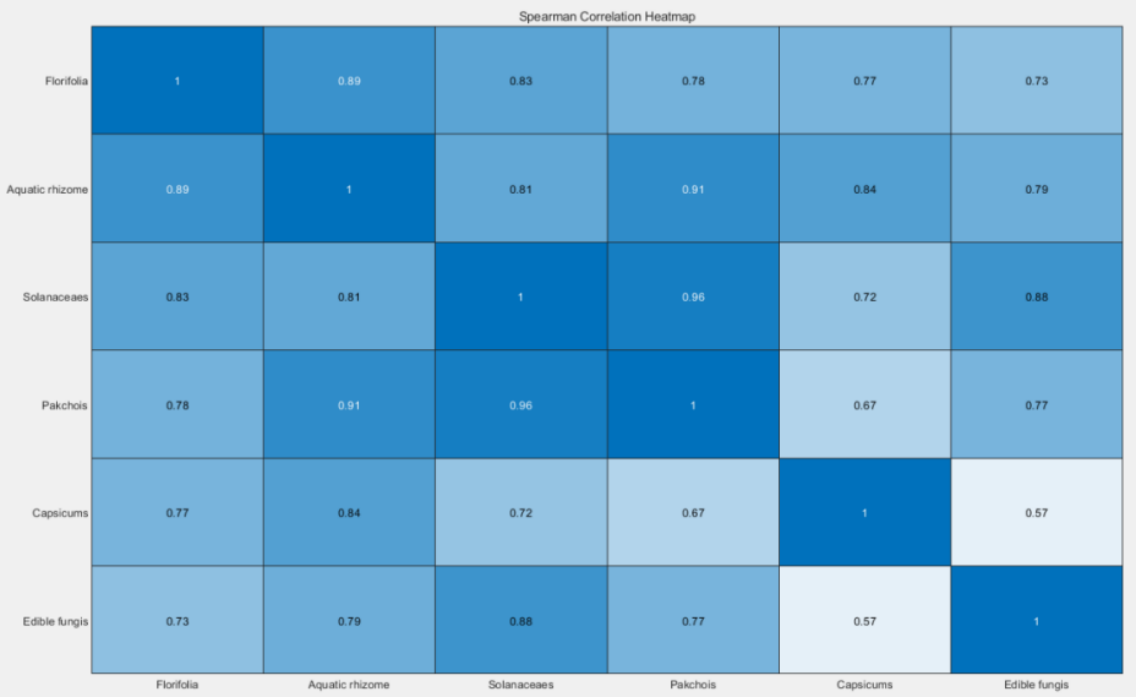

Fig. 2 Correlation analysis of the categories

By classifying the study data. Integrate vegetable categories and corresponding total sales volume, vegetable single product and corresponding total sales volume. Then, according to the integrated data, the single product name and classification name are classification variables, and the total sales volume is quantitative variables. The variables were unified, and the classified variables were transformed into qualitative variables. In order to prevent the impact of the dimension on the subsequent results, the total sales volume is normalized, and the method of extreme difference value standardization is adopted, that is, the data pretreatment and the comparison of the total sales volume of a single product as shown in Fig 3 for data visualization.

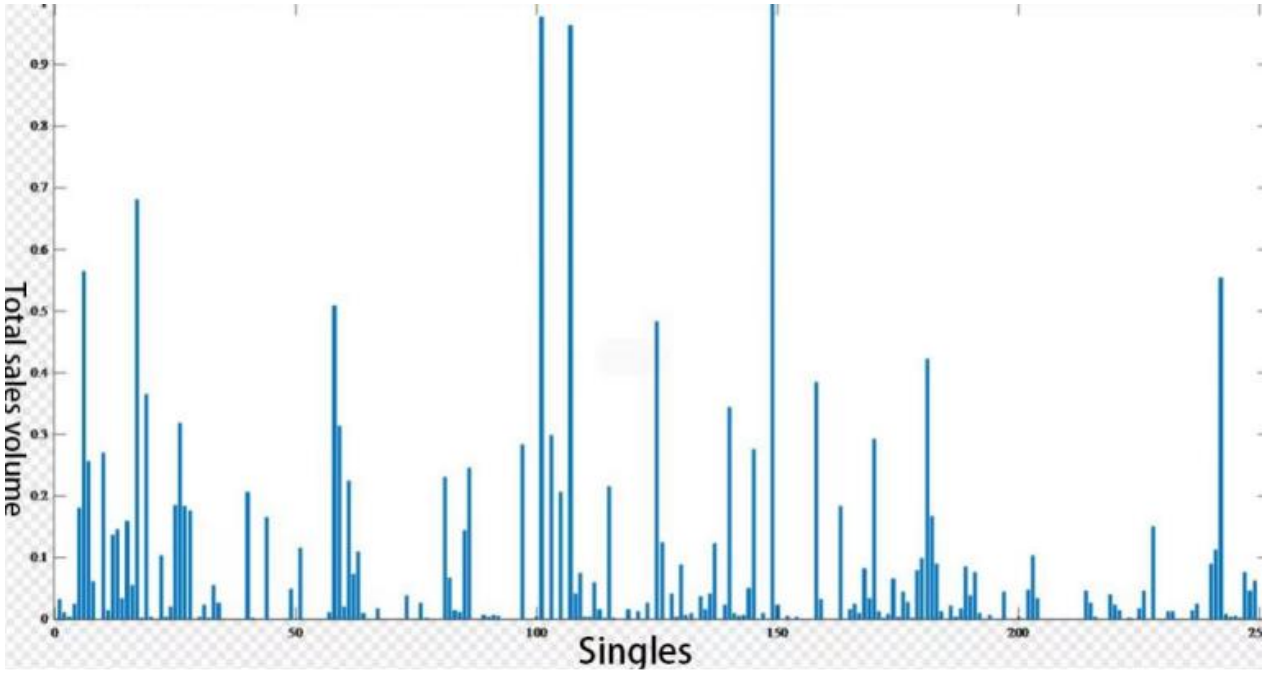

Fig. 3 Bar chart of the total sales volume of a single product

Secondly, in order to explore the correlation (or correlation coefficient) of different types of variables, the scatter plot, correlation relationship or correlation matrix can be considered to describe their correlation relationship. Correlation analysis was used to study the correlation between the two

variables, which was measured by the correlation coefficient. There are two types of correlation coefficient, namely, pearson correlation coefficient and Spearman correlation coefficient. Choose to use Matl ab to draw the judgment, use the scatter diagram to judge whether the classification name and the total sales volume conform to the linear relationship, normal distribution, use Q-Q, graph to check whether the data conform to the normal distribution, and then construct Spearman correlation coefficient, and finally get the specific correlation relationship. Taking the category and total sales volume as an example, bring the data into the Matlab to make a matrix scatter plot of each influencing factor as shown in Fig 4.

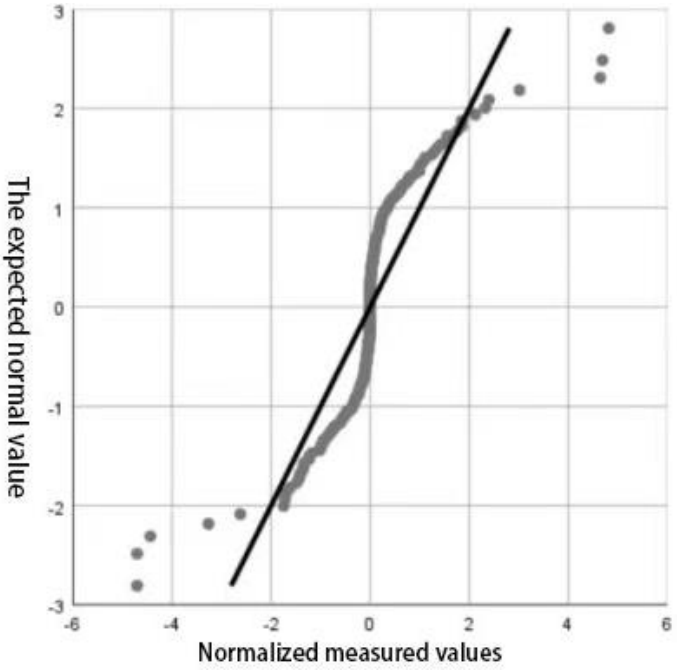

Fig. 4 Matrix scatter plot of category and sales volume

After obtaining the matrix scatter diagram of category and sales volume, it can be directly concluded that the factor correlation has no linear relationship. Therefore, the analysis of the normal Q-Q plots was performed.

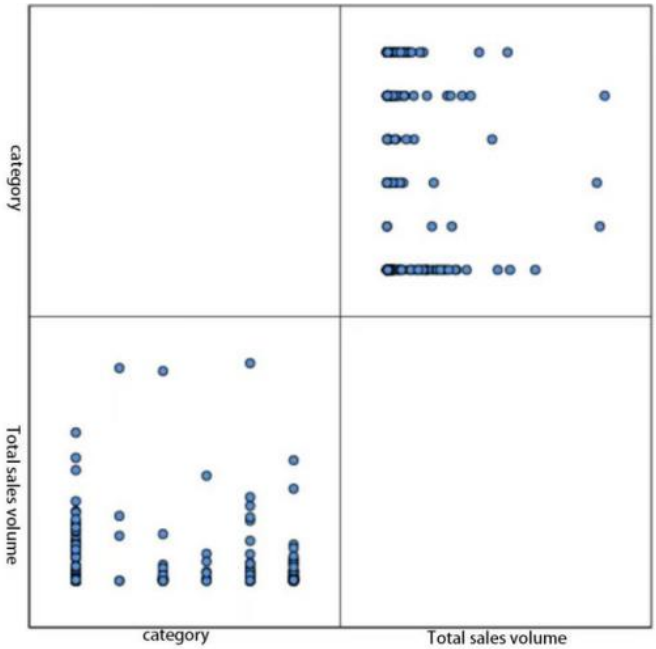

Fig. 5 Normal Q-Q plots in Fig

As shown in Fig 5, the point of this factor deviates from the line, so it does not conform to the normal distribution. Later, Matlab software was used to analyze Spearman's spearman correlation coefficient. During the calculation process, it was found that the P-value was 0.8953 and greater than 0.05, and there was no significant relationship between category and total sales volume. In conclusion, in this paper, the data is not applicable to Spearman correlation coefficient analysis, so we turn to cluster analysis, which is, all the data constitute similar groups. These similar groups are called clusters, which the complex data to standardize the complex structure data. After considering multiple conditions, the hierarchical-based clustering method is used to take each category, single product and sales volume of vegetables as indicators for hierarchical decomposition. By clustering, the cluster grouping is obtained, and the correlation of the same group is strong. The pedigree map was obtained as shown in Fig 6.

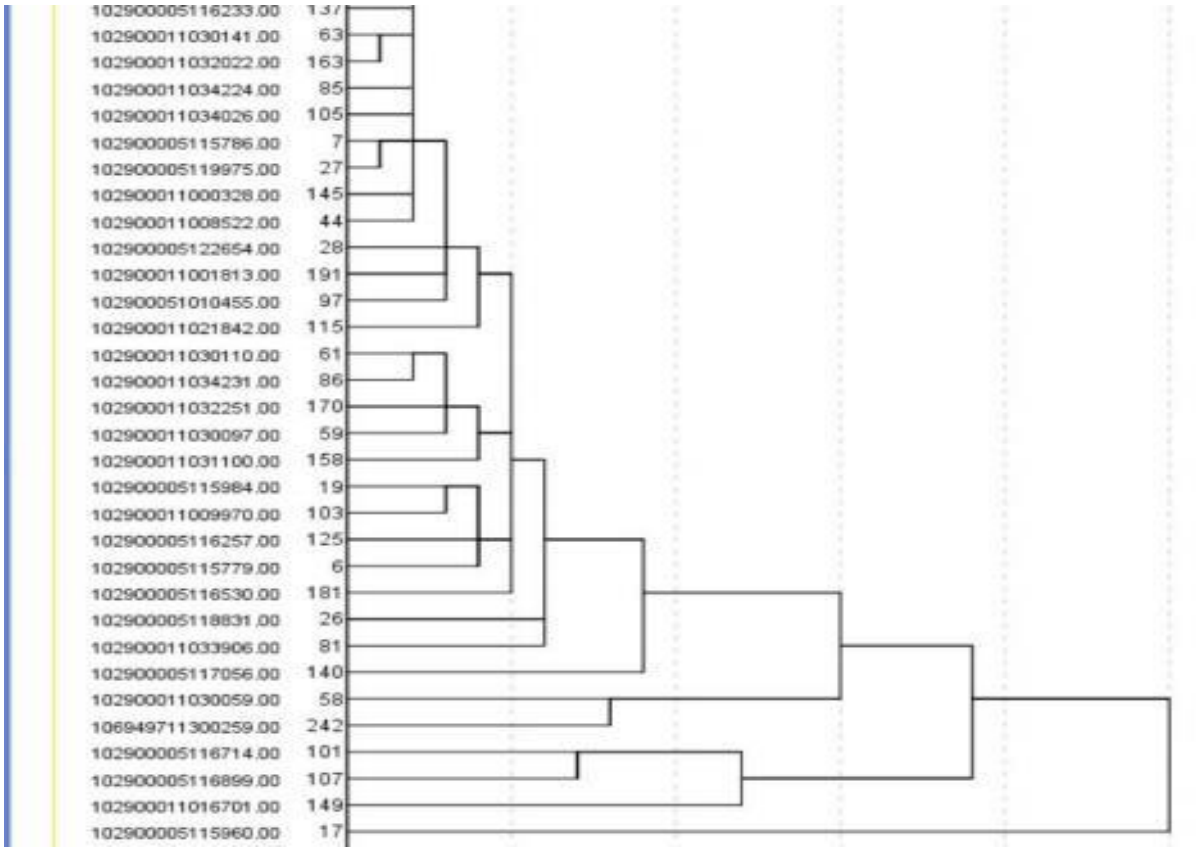

Fig. 6 Hierarchical diagram

Take the three categories, for example, As shown in Figure Figure 66, Make a longitudinal line (red line) across the three lines (black line), The numbers represented by the three black lines are in one class, for example, The Chinese cabbage itself, broccoli, Clean lotus root (1), Wuhu green pepper (1), of the same class, Other categories, According to this classification, Can be summed up into the three main categories, first kind, Family common dishes (like Chinese cabbage), Such sales and total categories are very large, It shows that common people use, Huge demand, Therefore, it is divided into common cooking dishes; second kind, Ordinary side dishes, It is found that these dishes are used as many side dishes and participate in a wide variety of dishes, This kind of dish is very much associated with the frequency of use; The third category, Other dishes, Found that the difference between the utilization rate and the sales volume of this type of dishes and the first two types of dishes is weak, So it belongs to the other dishes. Taking the above analysis as an example, you can choose several categories of analysis. Theoretically, the more categories, the more detailed the points, the better the relationship between them.

3. Explore the analysis rule of vegetable category and single product sales volume

3.1. Establish the supermarket revenue maximization model

In this paper, based on the analysis of the analysis law and mutual relationship of vegetable categories and individual product sales, the paper predicts the daily replenishment volume and pricing strategy of the supermarket in the next week (July 1-7, 2023), so as to maximize the profit of the supermarket[7].

On the basis of the analysis rule and mutual relationship of the vegetable category and single product sales volume, this paper predicts the daily replenishment volume and pricing strategy in the next week (July 1-7,2023), so that the supermarket can reach the maximum revenue[7

The curve over time obtains the change of sales volume in one day, calculates the products sold within one day, establishes a logistic regression model, and then makes a nonlinear fitting experiment on them. Logical functions are regression models for classification problems. Then develop the pricing strategy based on the existing rules[8][8]. In order to improve the model makes super yield maximum accuracy, after hundred experiments to find the shortcomings of strategy, and to optimize it, gives the optimized strategy, and for the next week (July 1-7,2023) daily replenishment and pricing strategy forecast, using ARIMA model, but considering the vegetable commodity sales have certain seasonal and cyclical, therefore, the season index forecast model is more representative[9].

In MATLAB, with code for ARIMA model compilation into data, running code, working matrix is the current available vegetables category data, including the month, weeks, average sales, average sales price, average wholesale price corresponding food information see the matrix Num 1, Num 2, matrix C recorded the corresponding ARIMA structure parameters, constant is C, AR parameters is the second order, MA parameters is the first order.

$$x_t = -5.8935 - 0.5177 \cdot x_{t-1} - 0.0199 \cdot x_{t-2} + 1 \cdot \varepsilon_{t-1} + \varepsilon_t \quad (2)$$

1.AR model (autoregressive model)

$$Xt = c + \varphi 1Xt - 1 + \varphi 2Xt - 2 + \cdots + \varphi \varepsilon t, \quad (3)$$

Among them, $\varphi 1, \ldots, \varphi\ \varepsilon t$ for its parameters

2.MA model (moving average model)

$$Xt = c\ + \varepsilon t + \theta 1\varepsilon t - 1 + \theta 2\varepsilon t - 2 + \cdots + \theta q\ \varepsilon t - q\theta 1, \ldots, \theta q \quad (4)$$

Where, is the moving average parameter.

3. ARMA model (autoregressive moving average model)

$$Xt = c + \varphi 1Xt - 1 + \cdots + \varphi\ \varepsilon t + \ \theta 1\varepsilon t - 1 + \cdots + \theta q\ \varepsilon t - q \quad (5)$$

After the difference of d times,

$$\Delta dXt = c + \varphi 1\Delta dXt - 1 + \cdots + \varphi\ \ + \varepsilon t + \theta 1\varepsilon t - 1 + \cdots + \theta q\ \varepsilon t - q \quad (6)$$

Gets the time series for the next week, and outputs the predicted results in a matrix,Among them, the first column is the average sales volume from July 1-7,2023, the second column is the average price, and the third column is the average wholesale price. For the calculation of the pricing model, we need to construct the demand price elasticity model, where P is the price and Q is the demand. According to the relationship between its elasticity coefficient and 1, it reflects the sensitivity of market demand to price changes. On the basis of the price elasticity of demand, assuming that the purchase quantity is an independent variable, remove the loss rate, and get the sales volume of vegetable goods. Compare the sales volume of vegetable goods at this time and the range of sales volume of the previous moment, calculate the range of price change, and thus continue to calculate the total income of the supermarket.

3.2. Analyze and formulate replenishment pricing strategies

Considering the replenishment plan of the supermarket based on the category, the relationship between the total sales volume of each vegetable category and the cost plus pricing was analyzed, and the total daily replenishment volume and pricing strategy of each vegetable category in the next week (July 1-7, 2023) was given to maximize the profit of the supermarket [10]. The former analysis uses regression analysis to obtain the linear regression equation between total sales and cost plus pricing, while the latter is a time series prediction model, which uses the resume penalty function of the regression equation to bring into the LSTM time series analysis (long period neural network prediction).

First of all, the detailed data of sales flow is summarized and the wholesale price data was processed to get the total sales of each vegetable category. Then, a regression analysis was performed to obtain a linear regression equation between total sales and cost-plus pricing for the subsequent modeling and optimization,as shown in Fig 7.

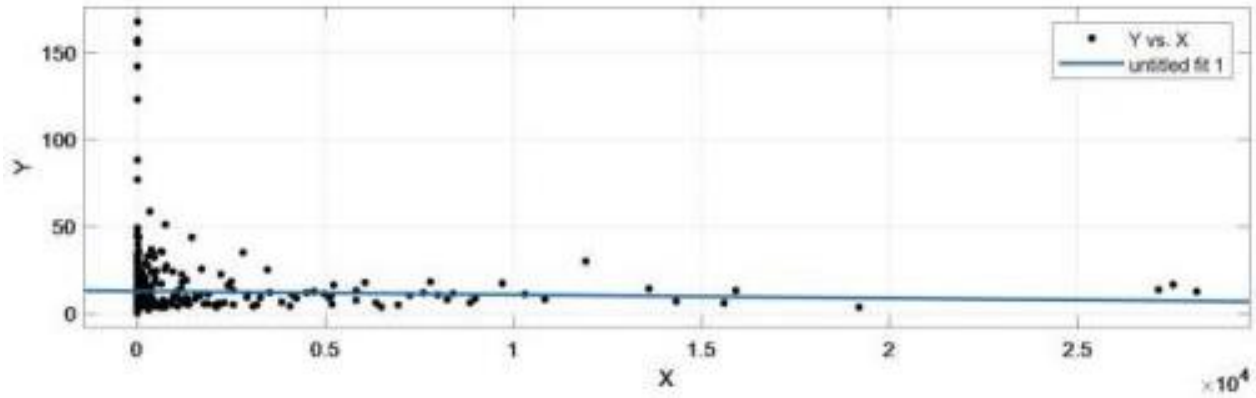

Fig. 7 regression analysis Fig

X means the total sales and Y means the average pricing, because the total sales is found to be zero, there will always be an average pricing, which complies with the law of market sales. Therefore, in order to prevent it from affecting the regression function, the case of X=0 is excluded in the regression, and then the regression is performed

$$f(x) = -0.0002013*x + 12.89 \quad (7)$$

The R-square: 0.9822 (the correlation coefficient between the measured data and reasoning data, close to 1 better) RMSE:3.01 (said the mean variance, the smaller the better) found the function is a minus function, but the degree of reduction is small, so it can be concluded that the vegetable category of sales and cost plus pricing is negative correlation, but not obvious, so although has the nature of the linear negative correlation but the influence is not big.

Finally, general time series predictions can be predicted using sales of each category in panel data such as seasonal ARIMA or exponential smoothing. Find the single product code and sales volume of different products, and with the time change curve, and get the sales change of a day. Add up the products sold within one day, build a logical regression model, and then do a nonlinear fitting experiment on them. Logical functions are regression models for classification problems. Use the common activation function sigmo id function, and then formulate the pricing strategy based on the existing rules. In order to improve the model makes the maximum accuracy, hundreds hundred experiments to find the shortcomings of strategy, and to optimize it, gives the optimized strategy, and for the next week (July 1-7,2023) daily replenishment and pricing strategy forecast, using ARIMA model, but considering the vegetable commodity sales has certain seasonal and cyclical, therefore, the season index forecast model is more representative.

In MATLAB, with code for ARIMA model compilation into data, running code, working matrix is the current available vegetables category data, including the month, weeks, average sales, average sales price, average wholesale price corresponding food information see the matrix Num 1, Num 2,

matrix C recorded the corresponding ARIMA structure parameters, constant is C,AR parameter is the second order, MA parameters is the first order.

Then, the pricing model is analyzed through the demand price elasticity model

$$E_t = \frac{\Delta P}{\Delta Q} = \frac{P_t - P_{t-1}}{Q_t - Q_{t-1}} \tag{8}$$

After this coefficient is obtained, assume that the purchase quantity is an independent variable, excluding the loss, which is the sales volume, and how the sales volume and the sales volume of the previous moment will change, so that the revenue can be calculated. The coefficient of each dish is shown in matrix E. Second, with the particle swarm algorithm for optimal quantity, forecast sales as the initial demand X0, before processing the data is the average sales, so when the revenue need *7, forecast sales price as the initial sales price S_0, wholesale price, set the quantity range (I set here is 0.5-1.5 times), quantity as a variable, minus the demand, demand relative X_0 change, calculate the new sales price S *, minus the wholesale cost benefits, through the particle swarm algorithm to find the maximum revenue,as shown in Fig 8 and Fig 9.

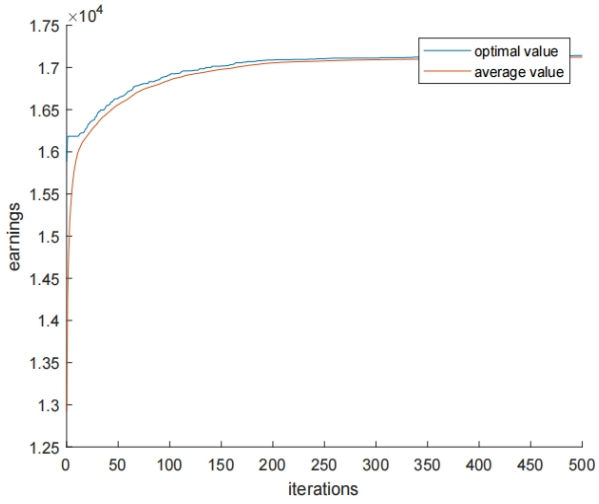

Fig. 8 Optimum and mean of returns under multiple tests

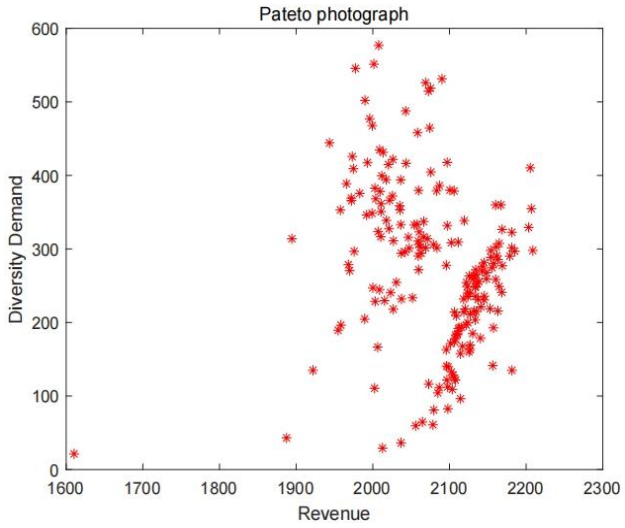

Fig. 9 Pateto plot of difference between return demand

4. Conclusion

This paper predicts the sales volume, sales price and wholesale price of each category of vegetables in the next 7 days, so as to give specific replenishment strategies.mosaic vegetables are the main selling vegetables of the supermarket, and the total sales of eggplant are the least. For aquatic root vegetables, the highest correlation was found with edible fungus vegetables, and roughly the same as with other vegetables. For cauliflower and Mosaic vegetables, the correlation between the two showed high similarity, but Mosaic vegetables showed more correlation to pepper vegetables. As for the pricing strategy, the dynamic price elasticity model is constructed by combining the relevant theory of the economics, and the ARIMA time series model is used to obtain the optimal commodity price, so that the supermarket profit is the highest, and the final profit is 1.1385e + 04. mosaic vegetables are the main selling vegetables of the supermarket, and the total sales of eggplant are the least. For aquatic root vegetables, the highest correlation was found with edible fungus vegetables, and roughly the same as with other vegetables. For cauliflower and Mosaic vegetables, the correlation between the two showed high similarity, but Mosaic vegetables showed more correlation to pepper vegetables.

Reference

[1] Li Yali. Consider the collaborative decision of joint replenishment and pricing for investment in preservation technology [D]. Chongqing: Chongqing Jiaotong University, 2022.

[2] Sun Yuping. Research on short-term demand forecast and inventory allocation of fresh commodities [D]. China University of Petroleum, 2021

[3] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO.(Version 1.0.11)[Online Application Software].Retrieved from https://www.spsspro.com .

[4] Zhong Yao, Nie Ying, Guo Jianxin, et al. Study on the characteristics and rules of vegetable price fluctuation in Beijing [J]. Price Theory and Practice, 2023 (6): 1-6.

[5] Jian-bo wang. Research on the management of fresh Goods in Beijing Yonghui Supermarket [D]. Beijing University of Chemical Technology, 2020

[6] Li Fangdi, Zhao Shangwen. Analysis of vegetable price trend in the first half of 2022 [J]. Agricultural Science and Technology and Information, 2023 (6): 162-165.

[7] Department of Trade and Foreign Economic and Statistics, National Bureau of Statistics. China Trade and Foreign Economic Statistical Yearbook [M]. Beijing: China Statistics Press, 2021. Department of Trade and External Economic Relations Statistics.China trade and external economic statistical yearbook-2021[M].Beijing:China Statistics Press,2021.

[8] AnKui. M Research on collaborative optimization of fresh supermarket supply chain inventory [D]. Guiyang: Guizhou University, 2021

[9] Lu Jing. Research on inventory control and dynamic pricing of fresh agricultural products [D]. Tianjin: Tianjin University, 2019.

[10] The compilation Committee of China Statistical Yearbook-2021. China Statistical Yearbook-2021 [M]. Beijing: China Statistics Publishing House, 2021. China Statistical Yearbook-2021 Editorial Board.China statistical yearbook-2021[M].Beijing:China Statistics Press, 2021.