Innovative Research on Pricing and Replenishment Decision Making of Vegetable Commodities Based on ARIMA Prediction Modeling

Haoxiang Cui *

School of Public Finance and Administration, Tianjin University of Finance and Economics, Tianjin, China

* Corresponding Author Email: 17622990317@163.com

Abstract. The pricing of vegetables in modern fresh produce superstores is generally based on the "cost-plus pricing" method, and reliable market demand analysis is particularly important for replenishment and pricing decisions. Against this background, this paper provides a complete set of vegetable merchandising strategies for supermarkets. Firstly, we detect and analyze the purchase data, study the relationship between the sales volume, visualize the sales volume data, and calculate the correlation between the sales volume of each category and individual product by using Pearson's algorithm and Spearman's algorithm respectively. Then, a pricing model was established, taking into account the cost of goods, expected sales volume and sales profit, and providing a daily replenishment total and pricing strategy for the coming week by analyzing the relationship between total sales volume and cost-plus pricing for each vegetable category. The model forecasts market demand to maximize revenue enhancement. Finally, the replenishment volume and the proposed pricing strategy for each individual item are determined using a portfolio optimization approach to ensure that the superstore achieves optimal profitability on a future day. The proposed model fits the actual needs, can reasonably solve the proposed problems, and is characterized by high practicality and algorithmic efficiency, which makes it suitable for all types of superstores to develop future sales strategies.

Keywords: Linear regression; Correlation analysis; ARMA time series modeling.

1. Introduction

In modern fresh food superstores, vegetables have a short freshness period and their quality will gradually deteriorate with the increase in sales time. In order to ensure that customers can purchase fresh vegetables, superstores usually replenish their stocks on a daily basis according to the historical sales and demand of each commodity. However, due to the fact that supermarkets sell many varieties of vegetables with different origins, and that vegetables are usually purchased from 3:00 a.m. to 4:00 a.m., merchants need to make replenishment decisions for each vegetable category on the same day without knowing exactly what the specific individual product and purchase price will be.

In order to balance costs and profits, the pricing of vegetables is generally "cost plus pricing" method [1-2]. For goods that are damaged or in poor condition, supermarkets will sell them at a discount to attract customers to buy them. In this process, reliable market demand analysis is particularly important for replenishment and pricing decisions.

From the demand side, there is often a correlation between the sales volume of vegetable items and the time of day. For example, during the evening hours, people may be more willing to purchase fresh vegetables for nutritional supplementation. From the supply side, the supply of vegetables is more abundant from April to October, but the limitation of sales space in supermarkets makes a reasonable sales mix extremely important. Therefore, superstores need to develop appropriate replenishment strategies and pricing strategies based on market demand and supply to ensure that customers can purchase fresh, high-quality vegetables to maximize revenue.

In this context, the problem of the thesis focuses on how to make replenishment decisions and pricing strategies for vegetables to maximize market demand and minimize losses. Therefore, in-depth data analysis and modeling is required to find out the best decision-making solution, which also needs to take into account several factors such as selling time, commodity characteristics and availability. This will help to improve the operational efficiency and customer satisfaction of fresh food supermarkets.

In order to solve the above problems, this paper provides a complete set of vegetable merchandising strategies for superstores to help them improve sales effectiveness, reduce inventory costs and increase customer satisfaction. First, the distribution patterns and interrelationships of the sales volume of each vegetable category and individual product are analyzed to explore the possible correlations between different categories or individual products. Then, analyze the relationship between the total sales volume of each vegetable category and cost-plus pricing, make replenishment plan by category, and give the total daily replenishment volume and pricing strategy for the coming week. Finally, further develop the replenishment strategy for individual items in the vegetable category to maximize the profitability of the superstore under the premise of trying to meet the market demand.

2. Analysis of the sales pattern and correlation of vegetables in supermarkets

First, analyze the sales trend of vegetables by category and individual product. Using Seaborn and Matplotlib libraries to create a sales trend graph as shown in Fig. 1, the date (time) of sales is used as the X-axis, the number of sales is used as the Y-axis, and vegetables with different classification names are represented in different colors to visualize the sales trend of different vegetable categories. It can be seen that in the category of flowers and leaves as well as chili peppers have high sales in some specific months, which are seasonal commodities, while the other categories are more stable, with less fluctuations and ups and downs.

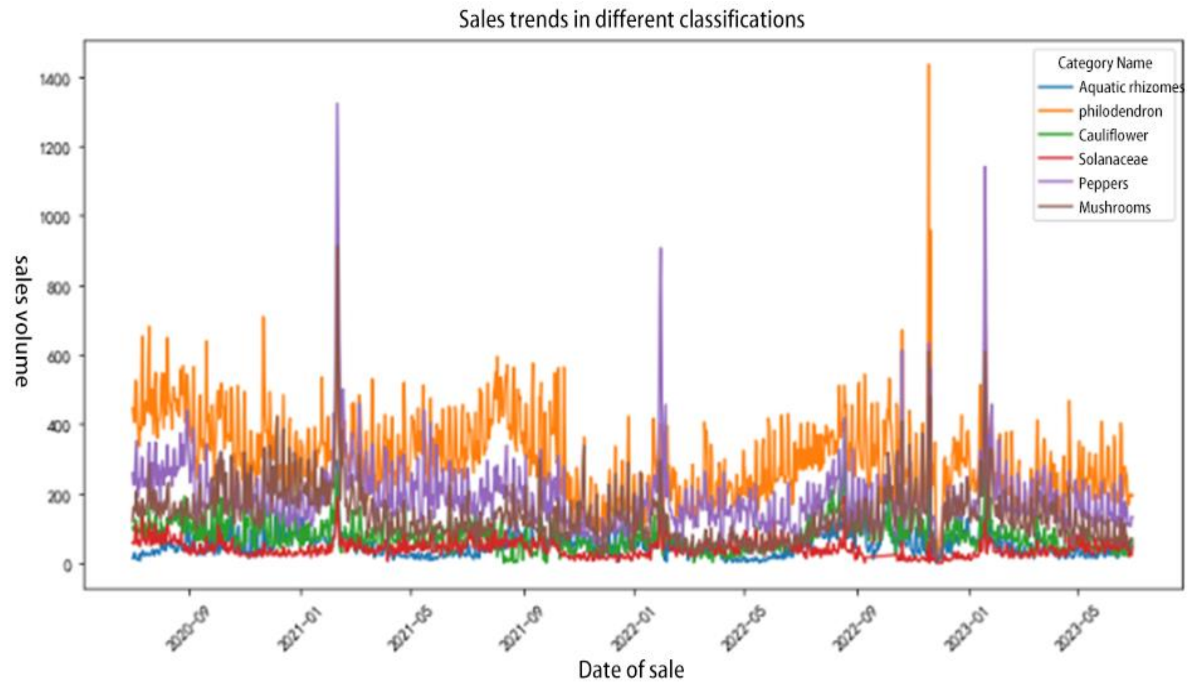

Fig. 1 Sales trend graph.

Then, analyze the total sales volume of vegetables for each category and individual item. Bar charts were created using Seaborn and Matplotlib libraries as shown in Fig. 2 to compare the total sales of vegetables with different category names in order to understand which categories have higher sales. In the past data, the different categories were dominated by foliage and chili peppers, while aquatic and eggplant were relatively few, judging that the reason was related to local eating habits; in the product subcategories, it can be seen that the individual items were mainly dominated by Wuhu shiangpei, broccoli, and Xixia shiitake mushrooms, and the others were generally the same in general.

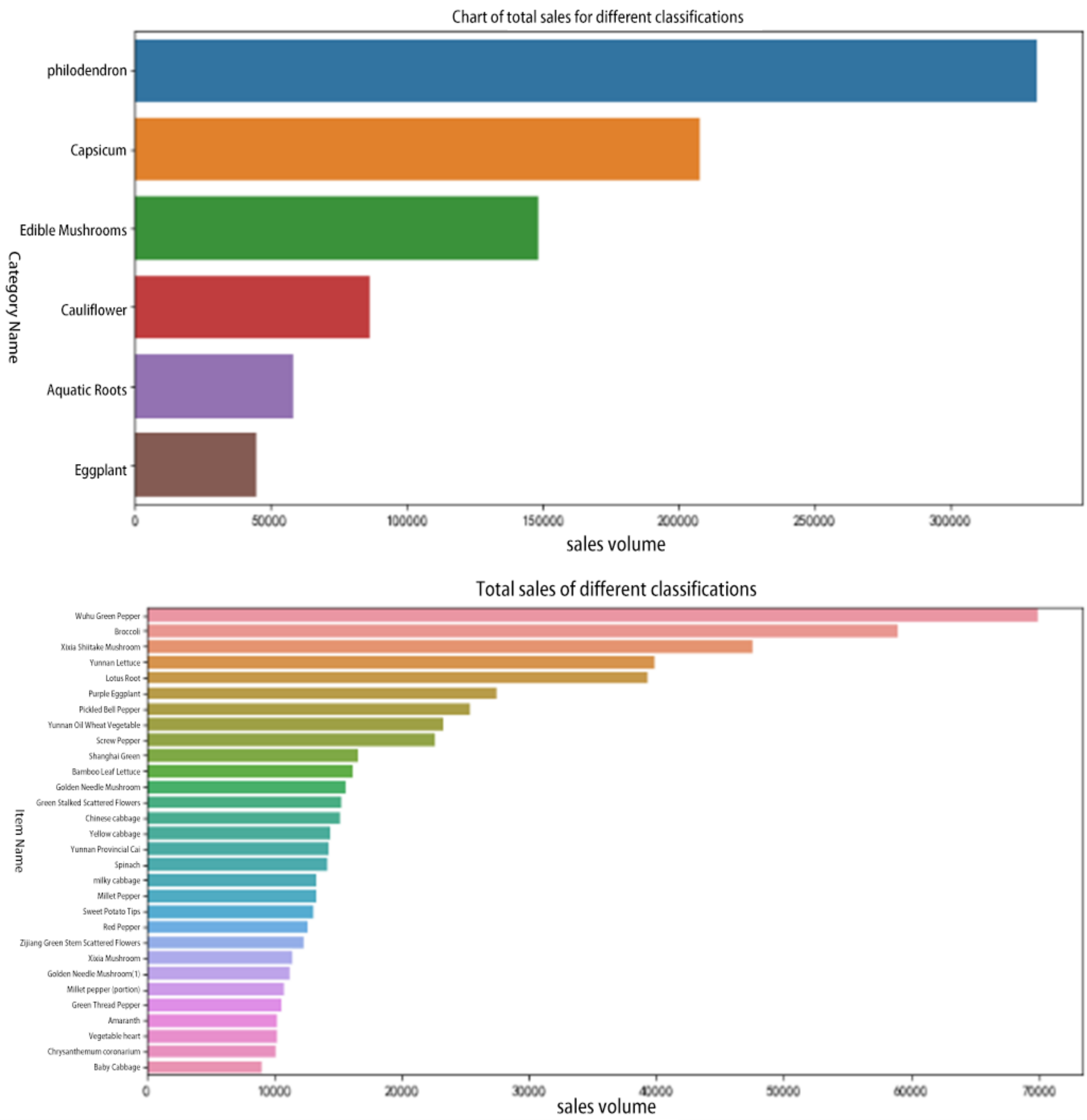

Fig. 2 Sales Bar Chart.

Finally, the sales volume correlation between different categorized names was calculated using the Pandas library. The correlation is presented visually using the Seaborn library to create a heat map as shown in Figure 3 to analyze the correlation (Pearson correlation coefficient) between different classification names [3]. The formula for Pearson correlation coefficient (r):

$$r = \frac{\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum\left(X_i - \bar{X}\right)^2 \sum\left(Y_i - \bar{Y}\right)^2}} \tag{1}$$

It can be seen that the correlation between chili peppers, aquatic roots, and edible mushrooms is relatively high, which correlates with local dietary habits.

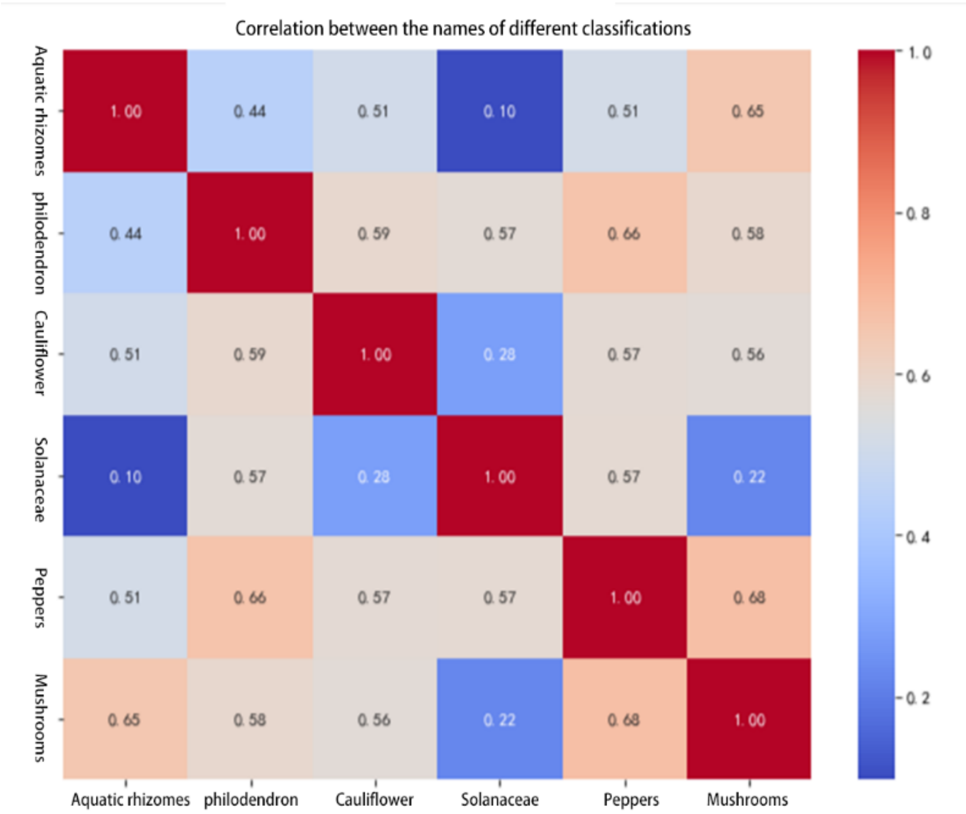

Fig. 3 Sales volume correlation chart.

3. Supermarket Pricing Strategy Development

3.1. Cost-plus pricing analysis

The total sales and cost plus pricing of each vegetable category were extracted from the collected data. The data was also cleaned and processed again to ensure the accuracy and completeness of the data. Then, the demand analysis was carried out. Calculate the average sales volume, average sales unit price, average wholesale price and average wastage rate for each vegetable category on different sales dates.

Calculate the cost of each vegetable category, which can include purchasing costs, transportation costs, and other related costs.

Determine the selling price for each vegetable category based on the cost-plus pricing methodology. Typically, the selling price can be set as a percentage of cost plus.

3.2. Regression Modeling

Linear regression is a statistical method used to model the relationship between one or more independent variables and the dependent variable. Here linear regression modeling is established for each vegetable category separately. Each vegetable category includes cauliflower, eggplant, foliage, pepper, edible mushrooms, and aquatic roots and tubers. For ease of description, they are collectively referred to as each vegetable category.

Multiple linear regression equations:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \tag{2}$$

The parameters of the model, which represent the strength of the relationship between sales volume and cost-plus pricing for each vegetable category, were estimated using the least squares method.

We assessed the accuracy of the model by calculating the mean square error (MSE) between the predicted and actual values. The results are shown in Fig. 4.

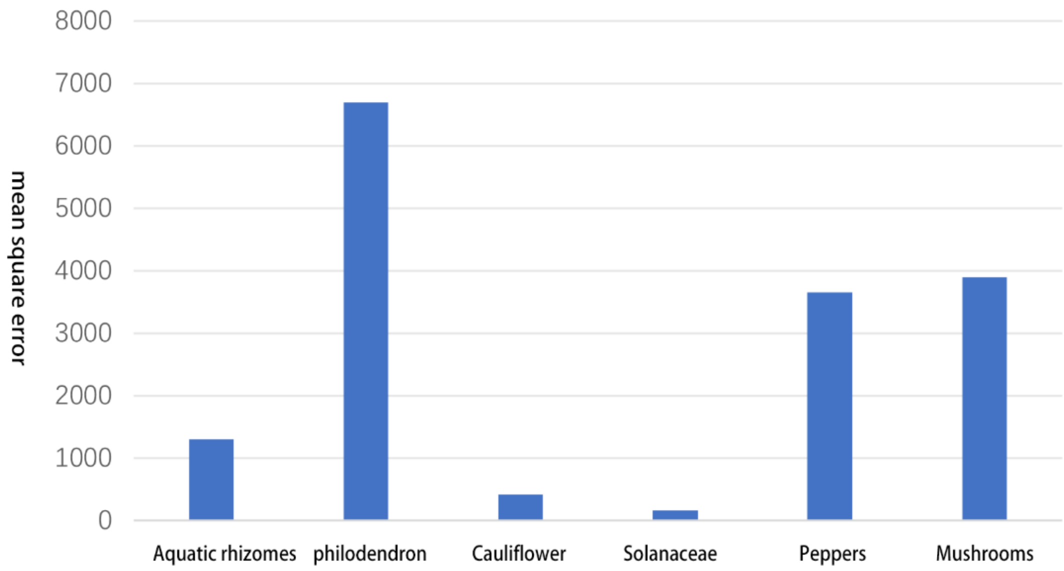

Fig. 4 Mean square error for each category.

3.3. ARIMA Time Series Modeling

Predicting the total daily replenishment for the coming week through simple linear regression still has some error. Due to the seasonal nature of prices, a time series model (ARIMA) can be used for forecasting [4]. Time series is a statistical analysis model that uses time series data to forecast future trends. It is based on the idea that the sequence of data formed from forecasts over time is considered to be a random sequence, which can be approximated using a model to describe the sequence. Once this sequence is determined, the model can predict future values based on the past and present values of the time series. In this model, we have attempted to predict the future unit price of a commodity based only on the unit price data up to the current month.

The ARIMA model consists of an autoregressive (AR) model and a moving average (MA) model. The AR model describes the relationship between the current and lagged values and predicts the future values using the historical data. The MA model looks at the future residuals using a linear combination of the past residual terms. The ARIMA prediction model can be written as follows.

$$\hat{p}^{(t)} = p_0 + \sum_{j=1}^{p} \gamma_j \, p^{(t-j)} + \sum_{j=1}^{q} \theta_j \, \varepsilon^{(t-j)} \tag{3}$$

Note; Here p is the order of the autoregressive model (AR), q is the order of the moving average model (AM), $\varepsilon^{(t)}$ is the error term between time, ranging between t and t-1, γ_j and θ_j are the fitting coefficients, and p_0 is a constant term.

Both ACF (Autocorrelation Function) and PACF (Partial Correlation Function) are functions that evaluate the linear relationship between the historical data and the current values.The ACF is given by:

$$ACF(q) = \frac{Cov(X_j, X_{j-q})}{Var(X_0)} = \frac{\frac{1}{n-q}\sum_{j=q+1}^{n}(x_j - \bar{x})(x_{j-q} - \bar{x})}{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2} \tag{4}$$

With the help of ARIMA model, we can make a simple prediction of the next day's price using past data. We can optimize the lattice strategy so that it can move the lattice according to the predictions given by ARIMA. The following figure depicts the effect of grid shift over time in the model. In our model, the lattice shift is determined by the weighted sum of the long term metrics from the MA and the short term metrics from the ARIMA model. The amount of grid shift is calculated by the following equation:

$$\overline{p_i}^{\{t\}} = p_i + \omega\left(MA^{\{t\}}(N) - MA^{\{t-1\}}(N)\right) + \mu\left(ARIMA^{\{t+1\}} - ARIMA^{\{t+1\}}\right) \tag{5}$$

The parameters ω and μ control the weights of the two metrics during the lattice movement. These two parameters will be adaptively adjusted during the backtesting phase, and we initialize ω to 0.3 and μ to 0.3 in the first cycle. Thus, the final decision is made by the adjusted lattice model to predict the price of the commodity.

3.4. Develop Daily Replenishment Totals and Pricing Strategies

In order to develop a pricing strategy, we need to consider the wholesale price of each category, the average wastage rate, and the expected sales volume. We will use a "cost-plus pricing" strategy that takes into account the cost of the item and the expected sales volume to determine an appropriate selling price.

Cost-based markup: Considering the cost of the item and the expected sales volume, we can determine an appropriate selling price based on the cost.

Considering the sales volume: Goods with higher expected sales volume can be set at a lower markup rate to increase sales, while goods with lower expected sales volume can be set at a higher markup rate to increase revenue. Based on the above considerations, we will set a markup rate for each category based on expected sales volume and cost to determine the daily pricing strategy for the coming week.

Cost-plus pricing strategy: We can determine the pricing strategy based on the cost of the item and the expected sales volume to maximize revenue.

After considering the above factors, do the calculations separately according to the following steps:

(1) Combine the wholesale price, wastage rate and projected sales volume in order to calculate the cost of each category. This is the cost (per kg) we have calculated for each vegetable category.

$$Cost\ (per\ kilogram)\ =\ wholesale\ price\ x\ (1\ +\ wastage\ rate/100) \tag{6}$$

(2) Determine the markup rate based on the expected sales volume: if the expected sales volume is < 20 kg, the markup rate = 150% (since these are low-volume items, we want to get a higher return per unit). If the expected sales volume is <50 kg, the markup rate = 130% (for medium volume items). If the expected sales volume is higher, markup rate = 120% (high volume items, where we want to increase sales volume through lower prices).

(3) Calculate the selling price for each category.

$$Selling\ price\ (per\ kg)\ =\ cost\ (per\ kg)\ x\ markup\ rate \tag{7}$$

Finally, the calculations are used to get the replenishment and pricing strategy for the super market vegetables.

4. Restocking Plan Development for Individual Products

Due to the limited sales space of vegetable items, in order to further develop the replenishment plan for individual items, to meet the total number of sellable individual items controlled within a certain range, and the order quantity of each individual item to meet the minimum display quantity requirements. Therefore, the original data is screened, and the data is first analyzed for 251 types of individual items.

Using the processed data, the relationship between replenishment and pricing for each type of vegetable individual item was calculated to derive a linear regression model as a constraint to the equation.

For the 251 types of vegetable individual items, different combinations of vegetables need to be selected on a daily basis. Summarized as:

$$q_i = \begin{cases} 0, \text{The ith individual item was not sold on that day} \\ 1, \text{The ith individual item is sold on the same day} \end{cases} \quad (8)$$

The total number of individual items control constraint can be expressed by summing the q_i switching variables.

For each item to be ordered to satisfy the minimum display quantity constraint, that is, it can be expressed as the total amount of each item e_i needs to be greater than the minimum display quantity.

For the calculation of profit, w_i is set as the profit margin. Therefore, the pricing of each individual item can be expressed as

$$y_i = (1 + w_i)c_i \quad (9)$$

Where, y_i denotes the pricing of the ith item, and c_i denotes the cost price of the ith item.

Thus the final total profit is expressed as

$$W_{tatol} = \sum_{i=1}^{251} w_i\, x_i \quad (10)$$

In summary, the optimization model can be constructed.

Since solving by general method has huge amount of computation, the model will be solved by using multi-objective particle swarm optimization algorithm [5]. Finally, the plan of single product is obtained.

5. Conclusions

This paper investigates replenishment decisions and pricing strategies for vegetables to help superstores increase sales effectiveness, reduce inventory costs and improve customer satisfaction. First, the purchase data are detected and analyzed, the sales relationship is studied in depth, and the sales data are visualized, and the Pearson algorithm and the Spearman algorithm are applied to calculate the correlation of the sales volume of each category and single product, respectively. Then, a reasonable pricing strategy is formulated and a pricing model is established, which takes into account the cost of goods, expected sales volume and sales profit, and provides a total daily replenishment and pricing strategy for future superstores by analyzing the relationship between the total sales volume and cost-plus pricing of each vegetable category. The model forecasts market demand to maximize revenue enhancement. Finally, the replenishment volume and proposed pricing strategy for each individual item are determined using a portfolio optimization approach to ensure that the superstore achieves optimal profitability at a future date. This requires consideration of expected profits, minimum display volume requirements, and other constraints. The proposed model fits the actual needs, can reasonably solve the proposed problems, is practical and algorithmically efficient, and is suitable for all types of superstores to develop future sales strategies.

References

[1] HAN Jun-Hua, GAN Sheng-Dao. A review of cost-plus pricing method [J]. Finance and Accounting Monthly, 2012(22):74-75.

[2] Wang Yu. Cost-plus pricing from KFC [J]. Journal of Economic Research, 2011(17):158-159+170.

[3] Gao Long. Research on small-lot material production demand forecasting model based on ARIMA [J]. Modern Information Technology, 2023, 7(15):97-101.DOI:10.19850/j.cnki.2096-4706.2023.15.021.

[4] Cheng Juanjuan. An empirical study on the relationship between research and teaching in colleges and universities--an analysis based on Pearson's correlation coefficient [J]. Science and Technology in Chinese Colleges and Universities, 2022(10):46-52.DOI:10.16209/j.cnki.cust.2022.10.016.

[5] Translated with www.DeepL.com/Translator (free version) YANG Wei,LI Qiqiang. A review of particle swarm optimization algorithms [J]. Chinese Engineering Science, 2004(05):87-94.