ISSN 2616-5902 Vol. 5, Issue 26: 164-171, DOI: 10.25236/AJBM.2023.052625

Replenishment and Pricing Strategies for Vegetable Commodities Based on Optimization Class Models

Yiwen Liu[1,*], Miaoru Li[1], Yifan Pu[1]

[1]*International Business College, Dongbei University of Finance and Economics, Dalian, China*
[*]*Corresponding author: 2691475791@qq.com*

Abstract: *Due to the nature of fresh commodities with short shelf life and strong seasonality, this paper is based on the study of the correlation between different vegetable categories and the relationship between sales volume and time, relying on historical sales data and using a variety of analytical methods to establish an optimization model of replenishment strategy and pricing strategy for fresh superstores. Firstly, the changes in the sales volume of goods in different categories with season or time are analyzed, the correlation between categories is quantified through Pearson correlation analysis to determine the purchase relevance of consumers between different categories, and the single product under the same category is analyzed by using systematic clustering method. Then, a one-dimensional linear regression model is established, and the maximum demand under the premise of satisfying the balance of supply and demand is used to determine the supply through the time series forecasting model, and one week is taken as the forecasting cycle, and the replenishment of each type of goods in the coming week is predicted by using Winters multiplier method and* $ARIMA(p,d,q)$ *model. Finally, the cost-plus pricing is analyzed to determine the single-item pricing optimization model, and the multi-objective particle swarm optimization algorithm is used to analyze and solve the pricing model for single-day replenishment strategy. The pricing model proposed in this paper can reasonably help the market to realize the balance of supply and demand, promote fair competition in the market, improve the market structure, and then realize the sustainable development of the economy. The model can be further generalized to other similar studies to help the pricing situation of other commodities in life.*

Keywords: *Commodity Pricing, Systematic Clustering, Pearson's Correlation Coefficient, Time Series Forecasting, Particle Swarm Optimization Algorithm*

1. Introduction

As of 2019, the scale of China's retail market has exceeded 5 trillion yuan, and since 2013 relies on the continued growth of more than 6%. At the same time, with the improvement of national disposable income and the development of planting technology, people's demand for fresh food categories has gradually become broader, and new requirements have been put forward for fresh food superstores in terms of service, product quality, shopping experience, etc. [1-4].

In the process of selling fresh food, the shelf life of commodities is generally shorter, with the increase in sales time, the quality of commodities will also show a substantial decrease, therefore, fresh food merchants need to analyze and understand the sales of different commodities and freshness of time to determine the time and quantity of replenishment. At the same time, in the process of transportation and sale of commodities, fresh commodities usually have a certain percentage of shipping loss phenomenon, which also poses a challenge to the pricing strategy of commodities. How to reduce the loss through scientific replenishment strategy and how to establish appropriate pricing strategy to make the merchants profitable are the problems to be solved and analyzed [5-8].

2. Model formulation and solving

2.1 Analysis of sales volume distribution patterns and correlations

First, taking each category as the object of study, the pattern of distribution of each category was first analyzed by descriptive statistics as shown in Figure 1. From a spatial point of view, by analyzing the overall sales of each category of vegetables during several years, the sales of flower and leafy vegetables topped each category, accounting for about 42.12% of the total sales, followed by chili peppers and edible

ISSN 2616-5902 Vol. 5, Issue 26: 164-171, DOI: 10.25236/AJBM.2023.052625

mushrooms, and eggplant sales were the least.

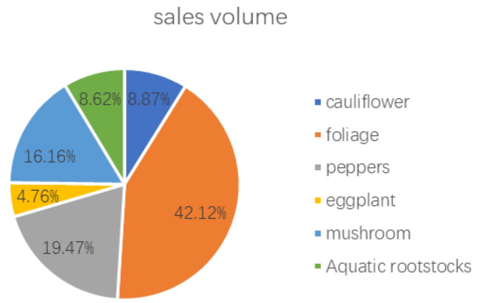


Figure 1: Pie chart of total sales by category

From a temporal perspective, there is volatility in the sales volume of different categories of goods in different quarters, with floriculture and foliage goods showing the greatest fluctuation, followed by edibles, and the total sales volume of the remaining categories of goods showing less variation from quarter to quarter, as shown in Figure 2.

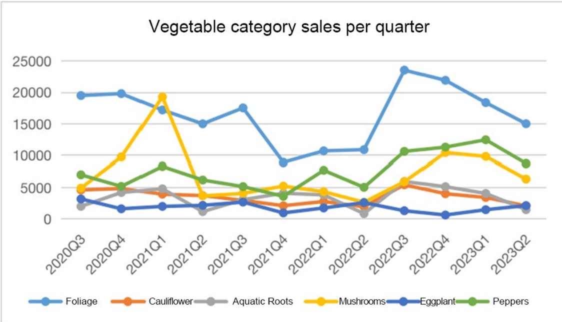


Figure 2: Line graph of category sales volume by quarter

An analysis of the sales of aquatic roots and tubers on a monthly basis is shown in Figure 3, where we find that the sales of goods in this category tend to be the lowest for the year from April to July, while sales gradually increase in the fall. Around February-March each year, there is a sudden increase in the sales of aquatic rhizomes, with the highest sales of the year.

As a representative product of aquatic root commodities, lotus root, its maturity time is basically in the fall of each year, so the supply in the fall is large, as a seasonal vegetable, by the consumer demand, and therefore present in the fall sales of large characteristics.

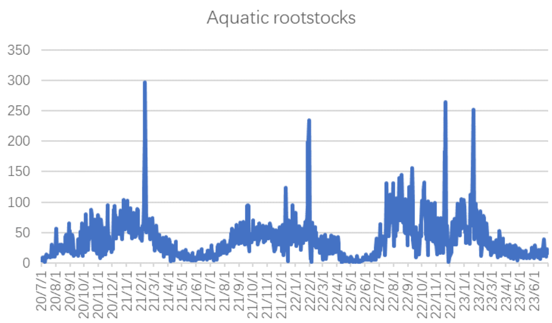


Figure 3: Monthly sales volume of aquatic root vegetables by individual product

In order to better analyze the correlation between the sales volume of each category of vegetables, so as to explore what kind of similar or opposite connection exists between certain vegetable categories in the sales process, this paper considers the Pearson correlation coefficient method to construct a model and carry out correlation analysis.

Pearson correlation analysis is a statistical method for measuring the strength of a linear relationship between two variables based on the Pearson correlation coefficient. It is calculated as follows.

ISSN 2616-5902 Vol. 5, Issue 26: 164-171, DOI: 10.25236/AJBM.2023.052625

$$\rho_{xy} = \frac{cov(X,Y)}{\sigma_x\sigma_y} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sqrt{\sigma_x^2\sigma_y^2}} \tag{1}$$

The Pearson's correlation coefficient of the sample can be obtained by estimating the covariance and standard deviation of the sample, which is often denoted by r. The expression for the Pearson's correlation coefficient is

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{X})(y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{Y})^2}} \tag{2}$$

In order to further obtain the correlation between the categories, this paper conducted a Pearson correlation analysis on the sales data of each category to determine the relationship between the categories, and plotted the heat matrix of Pearson correlation coefficients between the products as shown in Figure 4.

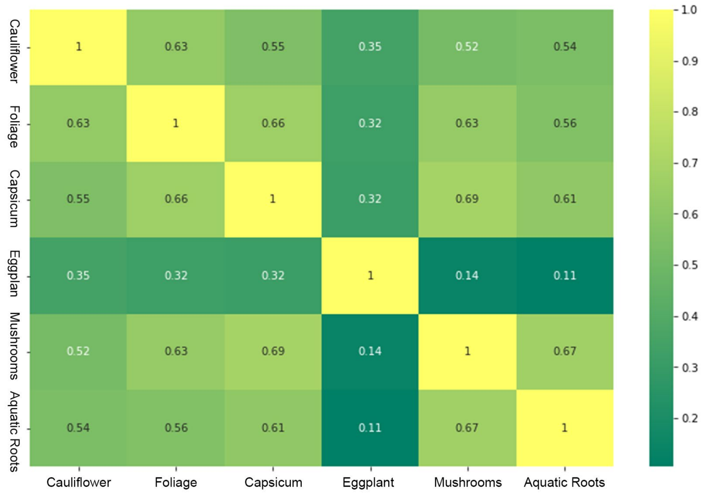


Figure 4: Heat matrix of Pearson correlation coefficients between vegetable categories

From Figure 4, the Pearson correlation coefficient between chili peppers and edible mushrooms is 0.69, which is the strongest linear relationship, and the linear relationship between edible mushrooms and aquatic roots and tubers is in the second place; the lowest degree of correlation is between eggplant and aquatic roots and tubers. From the results of the above analysis, it can be concluded that the fitness of chili pepper goods with other categories of goods is the strongest, and consumers mostly choose chili pepper goods to match with other categories of fresh food; eggplant has the lowest degree of correlation with other categories of goods, and consumers do not usually buy other goods to match with eggplant when they buy eggplant goods.

In order to illustrate the clustering results more representatively, this paper selects the chili pepper class to be classified to illustrate the use of average linkage method, using SPSS software clustering analysis is shown in Figure 5.

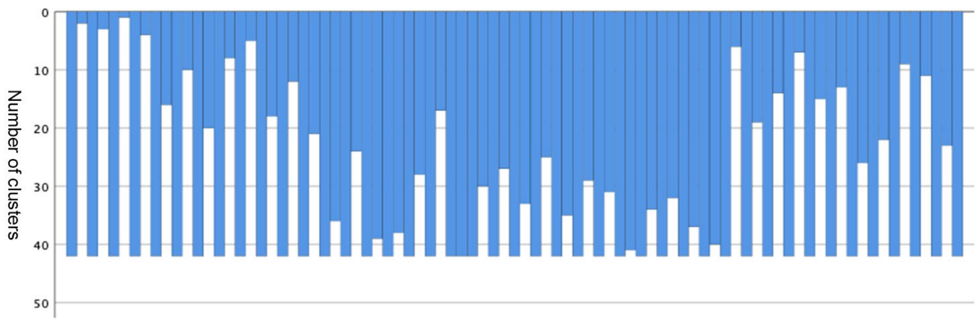


Figure 5: Clustered icicle diagram of vegetable dishes in the chili group

According to the above systematic cluster analysis icicle diagram, it can be seen that in the chili

ISSN 2616-5902 Vol. 5, Issue 26: 164-171, DOI: 10.25236/AJBM.2023.052625

pepper category, red pepper and green sharp pepper are the earliest to be aggregated, indicating that the two have a relatively strong correlation; red thread pepper and green hang pepper are the latest to be aggregated, and the correlation between the two is the weakest. Overall the clustering effect is not obvious, indicating that consumers have no obvious focus preference for the purchase of single products in the chili dishes.

2.2 Total daily replenishment and pricing strategy development for each category

To explore the relationship between the total sales of each category and cost-plus pricing, we must first clarify the principle of cost-plus pricing. Cost-plus pricing is the method of setting product prices according to the unit cost of the product plus a certain percentage of profit, and is also a common pricing method used by most companies, the basic arithmetic principle is:

$$P = C \times (1 + W) \tag{3}$$

$$C = \frac{(C_{fixed} + C_{variations})}{X} \tag{4}$$

$$w = \frac{P_{Sell} - P_{Purchase}}{P_{Purchase}} \times 100\% \tag{5}$$

The shipping loss and quality variation of the product will have a certain impact on its sales price, so when considering cost-plus pricing, the loss rate of the product should be taken into account, so as to obtain the historical target pricing of each product, the arithmetic formula is:

$$P_c = (P_{wholesale} + R_{loss}) - X - (1 + \frac{P_{sell} - P_{wholesale}}{P_{wholesale}}) \tag{6}$$

After obtaining the target pricing of individual items, the average of the target pricing of individual items is used as the cost-plus pricing of the category, as shown in Figure 6.

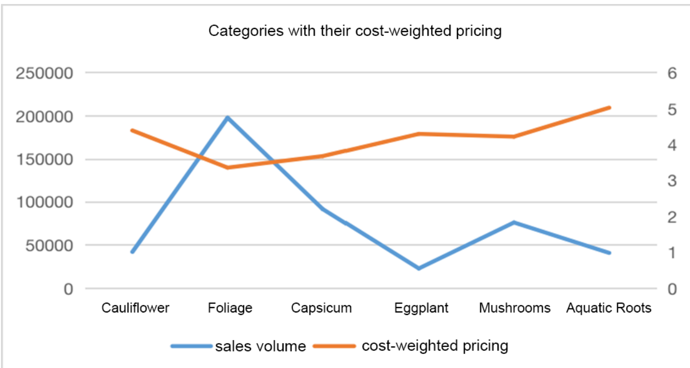


Figure 6: Line graph of each category with its cost-plus pricing

In order to further explore the relationship between total sales of each category and cost-plus pricing in depth, this paper establishes a one-way linear regression model, with total sales as the dependent variable y and cost-plus pricing of each category as the independent variable x, and sets up a one-way linear regression equation as:

$$y = \beta_0 + \beta_1 x \tag{7}$$

where β_0 is the constant term coefficient, and β_1 is the coefficient of the independent variable, the collated data were poured into SPSS software for data processing, and one-way linear regression was performed with its own linear regression function. The overall regression coefficient is not 0, there is a regression relationship between the variables, and the significance P-values are at 0.028 and 0.049, indicating that the regression model has a significance level of 0.05 or less, and a good fit. It is also evident that the model has a good fit.

From the constant coefficients as well as the coefficients of the independent variables, we can obtain the relationship between total sales y and cost-plus pricing in the case of category-based pricing as:

ISSN 2616-5902 Vol. 5, Issue 26: 164-171, DOI: 10.25236/AJBM.2023.052625

$$y = -88602.628x + 448175.683 \tag{8}$$

For every unit increase in cost-plus pricing, its total sales volume decreases by 88,602.628 kilograms.

As agricultural products, the sales of vegetable categories are often affected by seasonality, and there is a close relationship between time and product sales. Based on the theory of supply and demand equilibrium in economics, merchants, i.e., the supply side, have to realize supply and demand equilibrium under the framework of a market economy, i.e., to make their supply equal to the demand. In this model, it is assumed that in an imperfectly competitive market environment, the optimal replenishment quantity of a merchant is equal to the market demand, and all of them are sold. Therefore, the problem can be transformed into predicting the merchant's sales volume for the coming week.

On the demand side side, it can be analyzed from the historical data that there is a correlation between the sales volume of each category of goods and time. The sales volume data of six types of vegetable categories with daily time changes are categorized and regarded as six sets of time series variables. In order to construct the time series forecasting model, this paper chooses to use the SPSS expert modeler, which can analyze the historical time series data, and select the time series forecasting model with the highest degree of fit by observing whether there is a trend of seasonal changes and so on.

In this paper, six sets of sales data of vegetable categories over time are substituted into the SPSS expert modeler respectively, and in order to be accurate to the day, one week is taken as the forecasting period, and the time series variables measured by week, day are set. According to different time series characteristics, respectively, Winters multiplicative (exponential smoothing method), ARIMA (p,d,q) model, etc. predicted the sales of various types of vegetables in the coming week, that is, for the merchants in the coming week to replenish the amount of goods.

$$\begin{cases} l_t = \alpha \dfrac{x_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1}) \text{ (Horizontal Smoothing Equation)} \\ b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \text{ (Trend Smoothing Equation)} \\ s_t = \gamma \dfrac{x_t}{l_{t-1} + b_{t-1}} + (1-\gamma)s_{t-m} \text{ (Seasonal Smoothing Equation)} \\ \hat{x}_{t+h} = (l_t + hb_t)s_{t+h-m(k+1)}, k = \left[\dfrac{h-1}{m}\right] \text{ (Predictive Equations)} \end{cases} \tag{9}$$

For data containing linear trends seasonal trends are unstable, the exponential smoothing method is used for forecasting, which is based on the following principle:

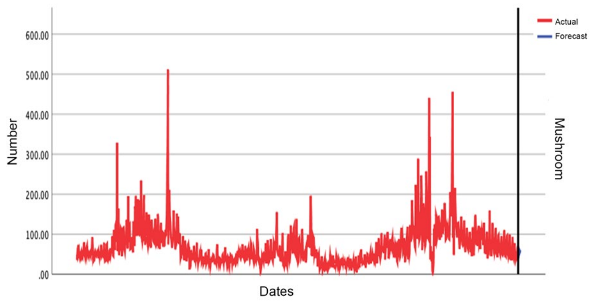
For smooth time series data, the ARIMA (p,d,q) model is used for forecasting, where the data are first differentiated and then modeled. Its forecasting principle is as follows:

$$y_t' = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}' + \varepsilon_t + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} \tag{10}$$

$$y_t' = \Delta^d y_t = (1-L)^d y_t \tag{11}$$

$$\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right)(1-L)^d y_t = \alpha_0 + \left(1 + \sum_{i=1}^{q} \beta_i L^i\right)\varepsilon_t \tag{12}$$

Constructing a time series forecasting model through SPSS expert modeler the following time series forecasting plot of sales for six dish categories is shown in Figure 7.



ISSN 2616-5902 Vol. 5, Issue 26: 164-171, DOI: 10.25236/AJBM.2023.052625

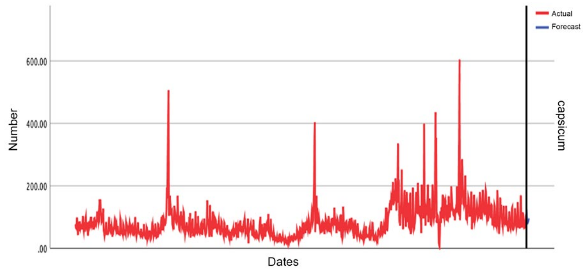


Figure 7: Time series forecast of future sales of chili and edible mushroom vegetables

The smooth R-square of the above time prediction models are all between 0.5 and 0.7 with a significance level of 0.000*, which indicates that the prediction models of this type are well fitted.

For the design of the pricing strategy for each vegetable category for the coming week, we use the eggplant category as an example to solve the analysis, and the rest of the categories show the final pricing results.

Set the decision variable to xi eggplant $i = 1,2,3,4,5,6,7$, denoting the sales volume of eggplant on day i. The decision variable is set to x_i.

Vegetable pricing in superstores uses cost-plus pricing. As an effective pricing strategy, cost-plus pricing is widely used in several industry sectors to determine the selling price of a product or service. The core idea is to combine costs with expected profits to determine the final selling price. Combining the basic formula of cost-plus pricing and taking into account the existence of wastage of dishes, let the daily profit margin be w_i Tomato

In turn, the pricing $y_{i\_eggplant}$ is expressed as $y_{i\_eggplant} = (1 + w)_{i\_eggplant} \times c_{i\_eggplant}$. In the equation, $c_{i\_eggplant}$ denotes the predicted cost of eggplant for the day.

The profitability of the eggplant class on day i can be expressed as

$$W_{i\_eggplant} = w_{i\_eggplant} \times x_{i\_eggplant} \tag{13}$$

The total profit of the superstore for the coming week is

$$W_{total} = \sum_{i=1}^{7} (W_{i\_eggplant} + W_{i\_philodendron} + W_{i\_cauliflower} + W_{i\_capsicum} + W_{i\_mushroom} + W_{i\_aquatic\ rhizomes}) \tag{14}$$

In summary, the pricing optimization model can be written as:

$$Max \sum_{i=1}^{7} \sum_{j=1}^{6} W_{ij} \tag{15}$$

List the equations and inequality constraints in terms of the relationship between the sales volume of each category and its selling price, and the principle of cost-plus pricing:

$$s.t. \begin{cases} y_{ij} = \beta_{j0} + \beta_{j1} x_{ij} \\ y_{ij} = (1 + w_{ij}) \times c_{ij} \\ W_{ij} = w_{ij} * x_{ij} \\ x_{ij}, y_{ij}, W_{ij} \geq 0 \\ w_{ij} \in R \\ i = 1,2,3,4,5,6,7, j = 1,2,3,4,5,6 \end{cases} \tag{16}$$

This is calculated to obtain the daily profit margin w_{ij}, which determines the price y_{ij}.

2.3 Daily replenishment and pricing strategy development for each individual product

Considering the limitations of the actual situation, the shelf placement of single products often has weight and quantity restrictions. On the basis of the established time series model as well as the

ISSN 2616-5902 Vol. 5, Issue 26: 164-171, DOI: 10.25236/AJBM.2023.052625

optimization model, the selected 30 types of single products are substituted into the model, and the constraints are further optimized to strengthen the constraint strength, and finally the improved optimization model is obtained.

Set the decision variable as x_i, $i = 1,2,3...30$ denote the pricing of the ith individual item. If the sales volume of a certain single item is too small to display it, the superstore will bear a large comparative display cost. Therefore, setting the minimum display quantity of an item to 2.5 kg can be expressed as follows: $e_i \geq 2.5, i = 1,2,3...30$

For the calculation of profitability, we follow the analysis of cost-plus pricing and set the profit margin as w_i, which in turn leads to the pricing of each individual product as y_i, denoted by:

$$y_i = (1 + w_i) \times c_i \tag{17}$$

where c_i denotes the predicted cost of the ith individual product.

Total earnings can be expressed as:

$$W_{total} = \sum_{i=1}^{30} W_i X_i \tag{18}$$

In summary, the single-item pricing optimization model can be written as.

$$MaxW_{total} = \sum_{i=1}^{30} W_i X_i \tag{19}$$

Constraints are written after adding certain conditions:

$$s.t.\begin{cases} y_i = (1 + w_i) * c_i \\ y_i = k_i x_i + b_i \\ e_i \geq 2.5 \\ x_i, c_i \geq 0 \\ y_i, w_i \in R \\ i = 1,2,3...30 \end{cases} \tag{20}$$

Considering the huge computing data of 30 objects, in order to simplify the operation, this paper adopts the multi-objective particle swarm optimization algorithm to solve the optimization model.

Particle Swarm Optimization is a stochastic search algorithm proposed by Dr. Eberhart and Dr. Kennedy originating from the study of birds foraging for food [9-11]. The PSO algorithm simulates the process of finding food in a flock of birds, where each bird in the flock is viewed as a particle of the population in the algorithm. Birds are in motion when looking for food, they keep approaching the food by changing the position and speed to which they fly until they find the food, and the principle of PSO is the same, the sharing of information by the individuals in the flock causes the motion of the whole flock to produce the evolution process from disorder to order in the problem solution space, so as to obtain the feasible solution of the problem. The specific formulation of particle swarm optimization algorithm is as follows:

$$V_i^d(t+1) = \omega V_i^d(t) + c_1 R_1(t)\left(pbest_i^d(t) - X_i^d(t)\right) + c_2 R_2(t)\left(gbest^d(t) - X_i^d(t)\right) \tag{21}$$

$$X_i^d(t+1) = X_i^d(t) + V_i^d(t+1) \tag{22}$$

where the optimal position path of individual particle is pbest and the optimal position path of the whole group of particles is gbest. R1(t) and R2(t) are random numbers between [0,1]. ω is the inertia weight, and its value is non-negative. The larger the value, the stronger the ability to find the optimal solution and the weaker the ability to find the optimal solution locally, and vice versa the opposite conclusion. ω decreases with iteration, in the pre iteration period, the particles have a larger update rate, and the swarm of particles can better develop the whole solution space. In the late iteration, as the inertia weight decreases, the updating speed of the particles also decreases, then the range of each particle's search for the optimal will also decrease gradually, and all the particles will only explore the neighborhood they are in.

The iterative formula for ω is shown below:

$$\omega = \omega_{max} + \frac{(\omega_{min} - \omega_{max})(t-1)}{Max\_iter^{\circ} - 1} \tag{23}$$

where Max_iter is the maximum number of iterations. The global optimal solution is obtained by

ISSN 2616-5902 Vol. 5, Issue 26: 164-171, DOI: 10.25236/AJBM.2023.052625

iteration.

3. Conclusions

Automatic pricing and replenishment decision-making requires fresh produce superstores to make decisions on whether to stock, how much to stock and how to price based on historical sales data and demand for each commodity, and after making a full understanding of each commodity's sales flow, wholesale price, wastage rate and other information. Based on the correlation between different vegetable categories and the relationship between sales volume and time, this paper relies on historical sales data and uses various analytical methods to establish an optimization model of replenishment strategy and pricing strategy for fresh produce superstores. Finally, the single-item pricing optimization model is determined, and the multi-objective particle swarm optimization algorithm is used to analyze and solve the pricing model of single-day replenishment strategy. The pricing model proposed in this paper can reasonably help the market to realize the balance between supply and demand, promote fair competition in the market, improve the market structure, and then realize the sustainable development of the economy.

References

[1] Li X, Zhou S. Research on the development problems of China's fresh food super retail industry. Research on Business Economy,2021(23):35-37.

[2] Wang C. Proof of the formula for calculating Spearman's coefficient. Journal of Yan'an University, 1997(01):73-75+77.

[3] Annikar E, Zulaikz M. Research on systematic clustering method and its application. Value Engineering, 2019,38(17):254-258.

[4] Xu T. Research on Commodity Demand Forecasting and Supplier Evaluation of Fresh Food Distributors. Xi'an University of Technology, 2023.

[5] Lu Z, Nie W, Chen L. Urban rail transit passenger flow prediction based on ARMA model. Henan Science, 2018,36(05):646-651.

[6] Li S. Design of DSM time-of-use tariff pricing based on PSO algorithm. Harbin Institute of Technology,2017.

[7] Wu H. Improved particle swarm optimization algorithm and applications. Hefei University of Technology, 2023.

[8] Mao K, Bao G, Xu C. Particle swarm optimization algorithm based on asymmetric learning factor adjustment. Computer Engineering,2010,36(19):182-184.

[9] Kennedy J, Eberhart R. Particle swarm optimization. Proceedings of the International Conference on Neural Networks. ieee, 1995

[10] Zhang Y, Hu X, Yang H. A green product pricing model under consumers' heterogeneous environmental preferences. Statistics and Decision Making,2017(06):40-45.

[11] Zhang J. Research on Pricing Model of Fresh Products Considering Returns in E-commerce Environment. Xi'an University of Electronic Science and Technology,2018.