Advances in Computer, Signals and Systems (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2023.071110
ISSN 2371-8838 Vol. 7 Num. 11

Supermarket Vegetable Commodities Based on TOPSIS-ARIMA Modeling Optimization Research on Replenishment and Pricing

Zhao Xin[1], Lu Wenjun[2], Jiang Manyi[1], Song Zijun[1]

[1]*Department of Economic Management and Law, South China University, Hengyang, 421001, China*

[2]*School of Resources, Environment and Safety Engineering, South China University, Hengyang, 421001, China*

Keywords: Autoregressive modeling, autoregressive integral sliding average modeling, cost-plus determination

Abstract: With the development of social economy, green and healthy food has gradually become the primary choice of consumers, thus intensifying the competition of vegetable commodities, resulting in the supply of vegetables sometimes exceeds the demand. However, since vegetables are characterized by a short freshness period and deterioration of dish quality, different factors affecting the sales and selling price of the commodities are considered comprehensively to meet the superstore to obtain the maximum return. Firstly, considering that the cost-plus pricing of vegetable commodities has a strong correlation with the discount price, transportation loss rate and storage time, the Topsis model is established to evaluate the different degrees of influence of the above factors, which results in the degrees of influence of the discount price, the transportation loss rate and the storage time on the cost-plus pricing of 21%, 42% and 37%, respectively. Secondly, we calculated the values of the above four indexes and obtained the linear fitting function between the total sales volume and the indexes, and concluded that the discount price is positively related to the total sales volume, with the maximum slope of 9.3218 and the minimum of 0.64; while the cost-plus pricing is negatively correlated with the total sales volume, with the minimum slope of -13.12 and the maximum slope of -0.944, which indicates that when the discount degree is bigger and the cost-plus pricing is lower, each vegetable category will be affected by the discount price and the cost-plus pricing. The lower the discount level and the lower the cost-plus pricing, the higher the sales volume of each vegetable category. Then the autoregressive model (AR) and autoregressive integral sliding average model (ARIMA) are used to fit the maximum value of interest to the sales price and sales volume of cauliflower and aquatic roots and tubers over time in three years to form a training set, and finally the daily replenishment total and pricing of each vegetable category in the coming week are predicted to give advice to the superstores on replenishment and pricing to maximize the revenue of the superstores.

1. Introduction

With the economic and social development, the competition in the commodity economy market is getting more and more intense. Consumers in the standard of living and consumption level in the pursuit of green health, which also intensified the competition of vegetable commodities. Vegetable sales also face many problems, vegetables have a certain life cycle, which also leads to its short freshness period, even if the businessmen to take freshness measures, but also can not guarantee that the vegetables are fresh for a long time [1]. With the vegetable commodities deterioration, not fresh, the sales volume will decline, or even stagnant [2]. At this time, merchants will take the price reduction promotion, which to a certain extent will reduce the profit [3]. Therefore, pricing and replenishment strategies are crucial for merchants to obtain high profits, and this paper will specify a reasonable program to make automatic pricing and replenishment decisions for vegetable products based on product categories and sales volume data.

The automatic pricing and replenishment decision of vegetable commodities involves various factors: supermarkets will replenish commodities every day according to the historical sales and demand of each commodity [4]; supermarkets will take into account the varieties and origins of commodities and the transaction time of purchase; supermarkets will take price cuts to promote the loss and poor quality of commodities, and generally adopt the cost-plus method of pricing; supermarkets will also take into account the supply and demand of the market. In this paper, the replenishment plan is based on the category as a unit, to study the relationship between cost-plus pricing and the total sales of each vegetable category, and to formulate the total daily replenishment and pricing strategy of each vegetable category in the coming week (July 1-7, 2023) to maximize the profit of the superstore.

2. Materials and methods

2.1 Data acquisition and pre-processing

The data comes from a survey of supermarkets and contains product information and sales flow details for six specific vegetable categories, wholesale prices of vegetable products, and recent wastage rates of vegetable products.

2.2.1 Data preprocessing

Because the amount of data is too large, the abnormal value of the data is tested and eliminated. If the sales unit price is x_n , sales for y_n ,Taking x as an example, the outliers are tested and eliminated; The t -test criterion is as follows:

Let $x_1, x_2, \ldots\ldots, x_n$ be a random sample from the normal population distribution $N(u,\sigma)$, and its mean is

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \; . \tag{1}$$

When the absolute value of $x_i - \bar{x}$ is greater than 0.02 (which can be transformed into a formula), it can be removed as a suspicious value x_d (α is one of $1, \ldots\ldots, n$) in this set of values, and then the average value is solved.

$$\overline{x_i} = \frac{1}{n-1}\sum_{i\neq\alpha} x_i \,. \tag{2}$$

Then use the *Bessel* formula to calculate the standard deviation

$$\overline{\sigma} = \sqrt{\frac{1}{n-2}\sum_{i\neq\alpha}\left(x_i - \overline{x}\right)^2}\,, \tag{3}$$

$$\overline{\sigma} = \sqrt{\frac{1}{n-2}\left[\sum_{i\neq\alpha} x_i^{\,2} - \frac{1}{n-1}\left(\sum_{i\neq\alpha} x_i\right)^2\right]}\,, \tag{4}$$

$$K(\alpha, n-2) = t(\alpha, n-2)\sqrt{n/(n-d)}. \tag{5}$$

Where $t(\alpha, n-2)$ is the $n-2$ distribution of t, the significance level is α, and $\left|x_\alpha - \overline{x}\right| \geq K(\alpha, n)\overline{\sigma}$.

Inspired by Reference [5], we select the significance level α as 0.01. If it satisfies Equation (4), the normal value is retained ; if it is not satisfied, further inspection is needed to eliminate suspicious values. Based on the elimination of the abnormal value of sales and sales unit price, the data is missing, and the Lagrange interpolation method is used to fill the data. If $(\mathrm{x}_i, \mathrm{z}_i)$ has n different data points, where $i = 0, 1, 2, 3, 4, \ldots\ldots, n-1$. Considering that sales are single-column values, rather than a set of data points, z_i can be determined as a fixed value. At this point, we choose $\mathrm{z}_i = 1$, the Lagrangian basis function $L_i(x)$ is

$$L_i(x) = \frac{\prod_{j\neq i}\left(x - \mathrm{x}_j\right)}{\prod_{j\neq i}\left(x_i - \mathrm{x}_j\right)}. \tag{6}$$

Through the above $L_i(x)$ function, Lagrange interpolation polynomial $L(x)$ is constructed.

$$L(x) = \sum_{i=0}^{n-1} y_i \cdot L_i(x). \tag{7}$$

By smoothing the data points, a curve is finally obtained to estimate other missing data points.

2.2 Methods to introduce

To study the relationship between the total sales volume of each vegetable category and the cost-plus pricing, the cost-plus data of each category should be collected first, which can be replaced by the sales unit price. In order to maximize the profits of the supermarket, it is necessary to define the indicators and explain the defined indicators, explain the quantitative indicators, and then substitute the collated data into the formula to establish a logistic regression model to obtain the relationship between the total sales volume of vegetables and time, and the relationship between the total sales volume and the unit price. By predicting the discount degree and total sales volume, the total profit is taken as the research object to establish a dynamic programming model, and finally the daily replenishment volume and pricing strategy of each vegetable category in the next week (July 1-7,2023) are obtained.

2.3 Model evaluation

Autoregressive model (AR) can capture the trend and periodicity of time series, analyze the autocorrelation and lag relationship in time series data, and predict the price and sales volume of different categories with high accuracy. At the same time, it can be used for various types of time series data, flexibly applied to different fields and problems, and adjusted and optimized as needed to reduce the error value.

The autoregressive integrated moving average model (ARIMA) can eliminate the non-stationarity of time series data by differential operation to meet the requirements of stationarity, so as to better model and predict.

The autoregressive model (AR) and the autoregressive integrated moving average model (ARIMA) have higher requirements for data, and need to have stability and correlation, otherwise the prediction results may be biased ; however, the autoregressive model (AR) has limitations on time series data with obvious seasonal and trend changes, and cannot accurately capture long-term trends and seasonal changes.

3. Model Establishment and Solution

3.1 Modelling

In order to establish the relationship between the total sales volume of each vegetable category and the cost markup pricing, some influencing factors should be quantified: the total daily sales volume of vegetable categories, the average unit price of vegetable categories, the average unit price of wholesale and the degree of discount. The details are as follows:

Let the total sales volume on day t be x_t, then

$$\mathrm{x}_t = \sum_{i=1}^{i=n} x_{i,t} - \sum_{i=1}^{i=n} y_{i,t} \,, \tag{8}$$

where $x_{i,t}$ is the daily sales volume of the tth vegetable on the ith day, n is the daily freight volume of the xth vegetable on the xth day, and x is the number of similar vegetables.

Let the average unit price under normal sales be p_t, then

$$p_t = \frac{\sum_{i=1}^{i=n} x'_{i,t} p_{i,t}}{\sum_{i=1}^{i=n} x'_{i,t}} \,, \tag{9}$$

where $x'_{i,t}$ is the normal sales volume of the tth vegetable on the ith day, and $p_{i,t}$ is the corresponding sales unit price.

The average unit price in the discounted sales state is $q_{i,t}$, then

$$q_t = \frac{\sum_{i=1}^{i=n} x''_{i,t} q_{i,t}}{\sum_{i=1}^{i=n} x''_{i,t}} \,, \tag{10}$$

where $x''_{i,t}$ is the discount sales volume of the t th vegetable on the i th day, and $q_{i,t}$ is the corresponding sales unit price.

The following calculation discount degree k_t, then

$$k_t = \frac{q_t}{p_t} \times 100\% . \tag{11}$$

Calculate the loss rate of vegetables:

$$s_t = \frac{\sum_{i=1}^{i=n} \left(x_{i,t} - y_{i,t} \right) s_{i,t}}{x_t} , \tag{12}$$

Where $s_{i,t}$ is the loss rate of various vegetables.

Let the average wholesale unit price of the category be u_t, then

$$u_t = \frac{\sum_{i=1}^{i=n} \left(x_{i,t} - y_{i,t} \right) u_{i,t}}{x_t} , \tag{13}$$

Where $u_{i,t}$ is the wholesale price of the i vegetable on the t day.

Through the quantitative solution of the above six vegetable category indicators, the linear fitting relationship between the total sales volume of single products and time, discount price, loss rate and markup pricing is established respectively. According to the analysis, it can be seen that time, discount price and loss rate also have a certain impact on markup pricing; in this paper, Topsis evaluation method is used to evaluate its indicators, and the weight distribution of each index of markup pricing is obtained.

Based on the above linear fitting equation, taking the fitting equation of time, discount price and loss rate to the sales volume of the category and the Topsis equation of markup pricing as the constraint conditions, the autoregressive model (AR) [6] is used to predict the sales volume and price of different types of items for one week. If

$$x_t = \Phi_0 + \Phi_1 x 1 + \Phi_2 x_{t-2} + \ldots + \Phi_p x_{t-p} + \varepsilon_t , \tag{14}$$

Where $\left| \Phi_p \right| < 1$ is the correlation coefficient of the sales volume of each different item, ε_t represents a white noise sequence.

From the long-term law of the sales volume of the previous items, the distribution functions of white Hypsizygus marmoreus (2), purple shellfish, Cordyceps sinensis and local Chinese cabbage are relatively concentrated and stable. At the same time, due to the stationary nature of the AR function [7], minus the $\mu = \dfrac{\Phi_0}{1 - \Phi_1 - \Phi_2 - \ldots - \Phi_p}$ of the sales volume of the single item, the distribution function of white Hypsizygus marmoreus (2), purple shellfish, Cordyceps sinensis and local Chinese cabbage is relatively concentrated and stable.

$$x_t - \mu = \Phi_1 \left(x_{t-1} - \mu \right) + \Phi_2 \left(x_{t-2} - \mu \right) + \ldots + \Phi_p \left(x_{t-p} - \mu \right) + \varepsilon_t , \tag{15}$$

Divide the $(x_t-\mu)$、$(x_{t-1}-\mu)$、…、 on both sides by the variance σ_0 of each item at the same time, we can get:

$$1=\Phi_1\rho_1+\Phi_2\rho_2+\ldots+\Phi_p\rho_p, \tag{16}$$

$$\rho_1=\Phi_1+\Phi_2\rho_2+\ldots+\Phi_p\rho_{p-1}, \tag{17}$$

$$\rho_2=\Phi_1\rho_1+\Phi_2+\Phi_3\rho_3+\ldots+\Phi_p\rho_{p-2}, \tag{18}$$

$$\rho_p=\Phi_1\rho_{p-1}+\Phi_2\rho_{p-2}+\Phi_3\rho_{p-3}+\ldots+\Phi_p. \tag{19}$$

Through the above relationship, the recursive relationship of various single products is summarized.

$$\rho_k=\Phi_1\rho_{k-1}+\Phi_2\rho_{k-2}+\ldots+\Phi_k\rho_{k-p}\left(k\geq p\right). \tag{20}$$

The above relationship is used to predict the sales volume and unit price of white Hypsizygus marmoreus (2), purple shellfish, Cordyceps flower and local Chinese cabbage after one week.

For cabbage moss, screw pepper and round eggplant (2) function distribution is relatively dispersed, the autoregressive integrated moving average model (ARIMA) is used for prediction analysis [8]. Firstly, the ADF test results of the above different categories are checked to judge the stability of the model. Then, according to the P value of Q statistics (P > 0.05), the model white noise is tested, and the information criterion AIC and BIC values are analyzed (the lower the better), and the model residual ACF / PACF diagram is analyzed. According to the model parameter table, the model formula is combined with the time series analysis diagram [9] for comprehensive analysis, and the backward prediction order result is obtained.

In order to maximize the profit of Shangchao, the profit of different single products on t day can be obtained by combining the above six quantitative indicators.

$$P=\sum_{i=1}^{i=6}\left[x_i(t)(1-k_i)p_i+x_i(t)k_iq_i-x_i(t)u_i\right]. \tag{21}$$

3.2 Model solving

Through (the calculation of the discount price formula), the following fitting function relationship can be obtained by using one-dimensional linear fitting (as shown in Table 1, 2, 3).

Table 1: The fitting relationship between the total sales of single products and the discount price

Single variety category	Fitting function relationship	R^2
Flower vegetables	y =1.417x+38.244	0.0226
Leaves and flowers	y=9.3218x+129.61	0.0226
Peppers	y=0.6477+66.739	0.0026
Solanaceae	y=0.64x+31.695	0.0035
Edible mushrooms	y=0.64x+31.695	0.1149
Aquatic rhizomes	y=0.8766x+24.261	0.0154

Where y is the total sales, x is the discount price.

Table 2: The fitting relationship between the total sales of single products and time

Single variety category	Fitting function relationship	R^2
Flower vegetables	$y=4E\text{-}05x^2-3.1629x+70821$	0.0588
Leaves and flowers	$y=0.0002x^2-20.527x+457473$	0.0575
Peppers	$y=0.0573x-2467.5$	0.1149
Solanaceae	$y=2E\text{-}05x^2-0.0361x+31.178$	0.0766
Edible mushrooms	$y=0.0001x^2-11.597x+258122$	0.0669
Aquatic rhizomes	$y=-4E\text{-}05x^2+3.5837x-79923$	0.0162

Where y is the total sales, x is the time.

Table 3: The fitting relationship between the total sales of single products and the loss rate

Single variety category	Fitting function relation	R^2
Flower vegetables	$y = 3.2152x + 4.8256$	0.0545
Leaves and flowers	$y = -12.355x + 339.09$	0.0796
Peppers	$y = 19.924x - 69.902$	0.0489
Solanaceae	$y = 1.9588x + 8.8911$	0.0094
Edible mushrooms	$y = -4.473x + 109.3$	0.0432
Aquatic rhizomes	$y = -2.5854x + 63.245$	0.1453

Where y is the total sales, x is the loss rate.

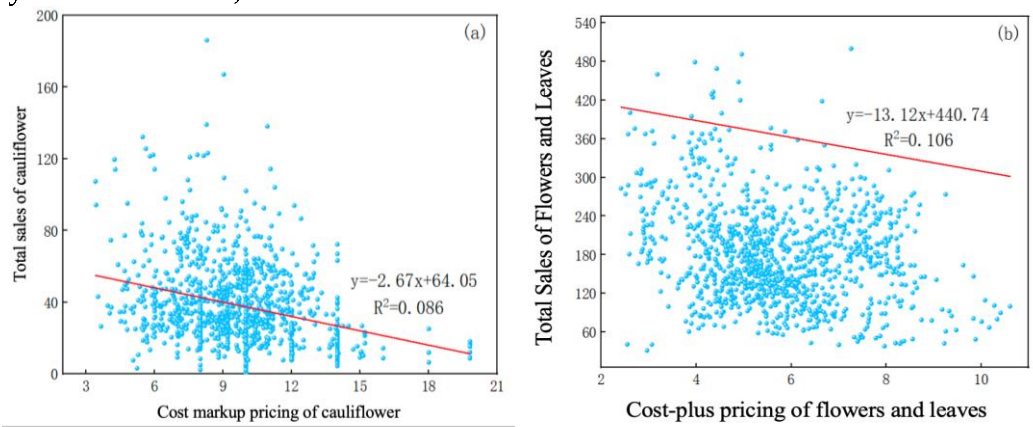

Figure 1: The relationship between the cost pricing of cauliflower and mosaic and their total sales respectively

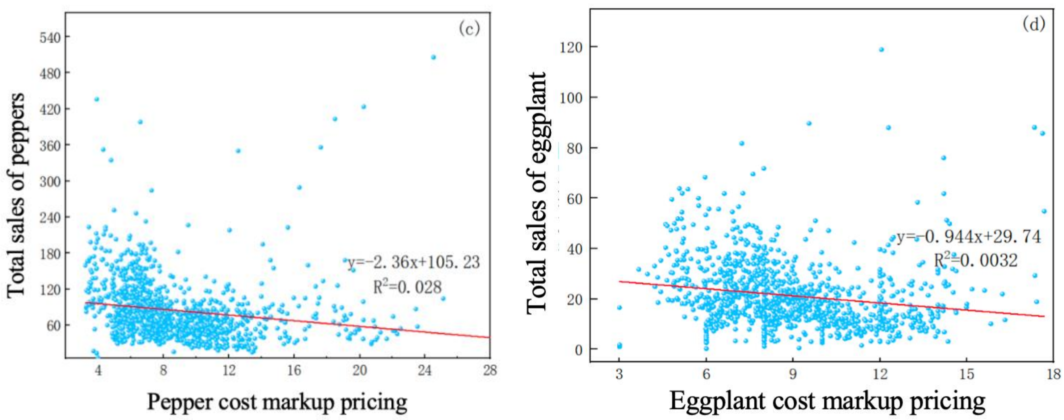

Figure 2: The relationship between the cost pricing of peppers and eggplants and their total sales volume

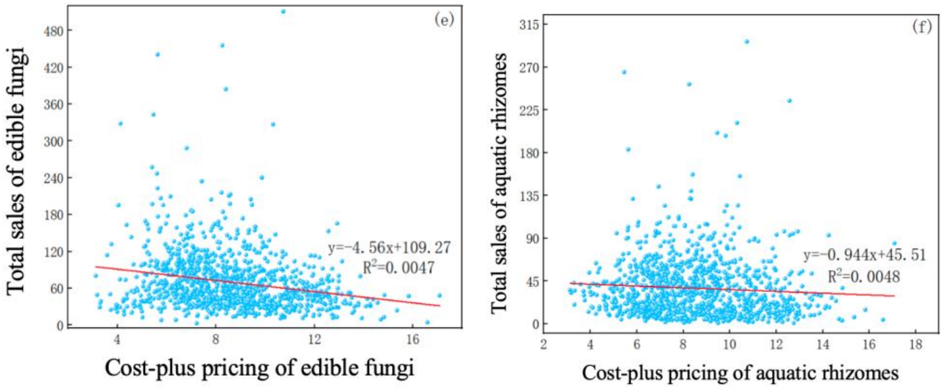

Figure 3: The relationship between the cost pricing of edible fungi and aquatic rhizomes and their total sales volume

Figure 1 shows the relationship between the cost pricing of cauliflowers and mosaics and their total sales. Figure 2 shows the relationship between the cost pricing of peppers and eggplants and their total sales. Figure 3 shows the relationship between the cost pricing of edible fungi and aquatic rhizomes and their total sales.

Considering that there are many data of time, discount price and loss rate, which may lead to large errors in the evaluation results, the relatively concentrated parts of the three data are extracted for Topsis evaluation, and the following relationship is obtained.

$$p_t = 0.21k_t + 0.42q_t + 0.37s_t \,. \tag{22}$$

After determining the functional relationship between markup pricing and the three, under certain constraints, the autoregressive model (AR) and autoregressive integrated moving average model (ARIMA) of the time series model are used to predict the sales volume and price of six different categories for the relative concentration and dispersion degree respectively. Considering that there are six different categories, we take cauliflower as an example to predict.

The constraint condition is

$$s.t.\begin{cases} y = 1.417x + 38.244 \\ y = 4E-05x^2 - 3.1629x + 70821 \\ y = 3.2152x + 4.8256 \\ p_t = 0.21k_t + 0.42q_t + 0.37s_t \end{cases}, \tag{23}$$

Table 4: Vegetable category forecast results in the next week

Category	Flower vegetables		Leaves and flowers		Peppers		Solanaceae		Edible mushrooms		Aquatic rhizomes	
	Selling price	The quantity sold	Selling price	The quantity sold	Selling price	The quantity sold	Selling price	The quantity sold	Selling price	The quantity sold	Selling price	The quantity sold
july 1	9.931	22.862	5.160	125.365	6.242	83.885	8.536	22.659	8.536	22.659	16.797	20.135
july 2	9.928	20.423	5.156	116.100	6.238	85.290	8.539	20.837	8.539	20.837	16.67	21.081
july 3	9.926	19.275	5.153	118.831	6.235	85.788	8.543	17.683	8.543	17.683	16.553	21.466
july 4	9.923	18.726	5.151	121.384	6.232	85.889	8.546	17.321	8.546	17.321	16.435	21.840
july 5	9.921	18.456	5.149	125.167	6.228	85.895	8.549	16.672	8.549	16.672	16.319	22.205
july 6	9.918	18.315	5.147	126.506	6.225	85.897	8.552	18.396	8.552	18.396	16.205	22.560
july 7	9.916	18.234	5.145	125.493	6.221	85.906	8.555	19.397	8.555	19.397	16.093	22.906

The constraint conditions of different categories are the fitting function relationship and Topsis function relationship of the three, and the commodity pricing of different categories is predicted [10].As shown in the figure, the prediction results of cauliflower, mosaic, pepper, eggplant, edible fungus and aquatic rhizome are respectively. Among them, the profit function is used to maximize the profit when training the function [11].

Through the above model training, the total daily replenishment and pricing of each vegetable category in the next week (July 1-7, 2023) are predicted to achieve the maximum profit of the supermarket (as shown in Table 4).

4. Conclusions

This paper deeply discusses the relationship between cost-plus pricing and discount price, loss rate and storage time, and analyzes the influence of these factors on pricing decision by establishing autoregressive (AR) and autoregressive integrated moving average model (ARIMA). Based on the research results of this paper, commercial supermarkets can obtain a clear strategic guidance on how to make decisions based on the predicted total replenishment and pricing to maximize profits. By referring to the linear fitting function of this paper, Shangchao can predict the future market demand more accurately, so as to formulate a more reasonable purchase and pricing strategy. This will help them maintain a leading position in the fierce market competition and obtain the largest commercial profit.

The research in this paper not only has practical application value and is of great significance to the commercial super industry, but also provides a new perspective and thinking for scholars and practitioners in the fields of economics, marketing and supply chain management. By understanding and using these models and forecasting methods, Supermarket and other related enterprises can better grasp market dynamics and make more scientific and effective business decisions.

References

[1] Yang Shuai, Huang Xiangmeng, Wang Junbin. Research on joint optimization strategy of fresh food shelf allocation and pricing [J].Supply chain management, 2022,3 (08):49-59.

[2] Cui Ligang, Li Yali, Liu Jinxing, etc. Joint decision-making on multi-product replenishment and pricing considering investment in preservation technology [J]. Industrial Engineering and Management, 2023, 28 (03) : 17-26.

[3] Yao Ziwei. Research on price decision of fresh products considering price and price reduction time [D].Southeast University, 2022.

[4] Wang Heng, Xu Qi. Research on retailer pricing and ordering strategy based on option contract under the correlation between market demand and sales price [J].Journal of Ningxia University (Natural Science Edition), 2017,38 (04) : 332-337.

[5] Yang Renti, Chen Wenzhang. When analyzing the test error, the t-test criterion is used to eliminate the bad value and its electrical algorithm [J].Sichuan Machinery, 1982 (05): 25-30.

[6] Abdelhakim A, Christian F. Two-stage weighted least squares estimator of the conditional mean of observation-driven time series models [J]. Journal of Econometrics, 2023,237(2PB).

[7] Wang Yan. Applied Time Series Analysis [M]. Beijing: Renmin University of China Press, 2005.

[8] Soumik R, Achal L, Pradeep M, et al. An ARIMA-LSTM model for predicting volatile agricultural price series with random forest technique[J]. Applied Soft Computing,2023,149(PA).

[9] Yanbo, Xu. Optimization model of rural supermarket location based on Entropy Weight TOPSIS[C]//International Conference on E-business & E-government. IEEE, 2011.DOI:10.1109/ICEBEG.2011.5881553.

[10] Özbilge Armağan, Hassini Elkafi, Parlar Mahmut. Optimal pricing and donation policy for fresh goods[J]. European Journal of Operational Research,2024,312(1).

[11] Jiang Y,Shang J,Liu Y, et al. Redesigning promotion strategy for e-commerce competitiveness through pricing and recommendation[J]. International Journal of Production Economics,2015,167.