Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/121/2025.20231

Automatic Pricing and Replenishment Decisions for Vegetable Products

Yuehan Fan[1,a], Dingding Chen[1,b],Yin Lu[1,c,*]

[1]*School of internet of things, Nanjing University of Posts and Telecommunications, Nanjing, China*
a. fyh2426137075@outlook.com, b. 1769157532@qq.com, c. luyin@njupt.edu.cn
**corresponding author*

Abstract: The perishable nature of vegetable products poses significant challenges in inventory management and pricing, often leading to high waste rates and economic losses. This study addresses these issues by proposing an automated pricing and replenishment decision model that leverages statistical analysis, regression methods, and time-series forecasting. Utilizing three years of sales data from a supermarket, the research develops models to analyze the relationship between sales volume and cost-plus pricing, optimize replenishment strategies, and predict pricing for various vegetable categories. The results demonstrate that dynamic pricing and replenishment decisions can reduce waste, increase profits, and enhance supply chain efficiency. The study highlights the practicality of using machine learning models like LSTM to forecast demand and pricing while acknowledging limitations such as simplifying assumptions and the need for multidimensional data. Future research directions include integrating external factors, such as climate impacts, to improve the model's adaptability and accuracy.

Keywords: Vegetable Inventory Management, Dynamic Pricing, Replenishment Decision Model, LSTM, Supply Chain Optimization

1. Introduction

In modern retail, vegetable products have become one of the most challenging areas in inventory management due to their perishable nature and demand volatility. Globally, food waste has become an increasingly severe issue. According to the Food and Agriculture Organization (FAO), approximately one-third of the world's food is wasted annually, with the waste rate for vegetables and fruits reaching as high as 40% to 50% [1]. This phenomenon not only results in resource wastage but also imposes significant environmental burdens and economic losses. In China, statistics from the Ministry of Agriculture indicate that about 20% of vegetables are damaged before reaching consumers each year due to improper preservation and related issues [2].

Managing vegetable products is particularly complex due to their short shelf life and fluctuating demand patterns. The limitations of shelf life make it difficult to store inventory for extended periods, and unsold products often result in expired waste. Simultaneously, the demand for vegetables fluctuates significantly due to factors such as seasons, weather, and holidays. Traditional pricing and restocking cycles often fail to adapt to these dynamic changes. Moreover, the lack of real-time information support in manual pricing and inventory management hampers operational efficiency.

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/121/2025.20231

Automated pricing and restocking systems, based on mathematical modeling and data analysis, offers great potential to significantly improve the management efficiency of vegetable products. By leveraging real-time data analysis and predictive models, these systems can more accurately adjust inventory and pricing. First, by analyzing historical data, the system can forecast future demand changes and avoid issues such as overstocking or insufficient stocking. For instance, by identifying the peak and off-peak seasons for certain vegetables, the system can increase inventory during peak seasons and reduce it during off-peak periods, thereby reducing waste. Dynamic pricing systems can flexibly adjust prices based on the remaining shelf life of vegetables and market supply-demand conditions, accelerating product turnover and reducing the risk of losses. For example, vegetables nearing the end of their shelf life can be discounted strategically to stimulate consumption and reduce waste. According to the research by Kunz et al., intelligent inventory management and dynamic pricing strategies can reduce food waste rates by 10% to 20% [3].

In response to these challenges, this paper proposes an automated decision-making system for pricing and restocking based on data analysis and mathematical modeling. By adopting a data-driven approach, the system aims to achieve scientific and efficient management of vegetable products. This study not only helps supermarkets and other retailers increase the sales profitability of vegetable products but also significantly reduces the wastage of perishable goods, thereby improving the overall efficiency of the supply chain. This paper first provides an overview of the relevant issues and then proposes a comprehensive solution based on statistical analysis, regression analysis, time series forecasting, and optimization models. Finally, the results are validated, and future research directions are suggested.

2. Methods

2.1. Model building preparation

We collected product information of six vegetable categories sold by a supermarket, as well as sales flow details and wholesale price data of each product from July 1, 2023 to June 30, 2023. In order to more intuitively represent the collected data, we drew a line chart and a bar chart based on the collected data:

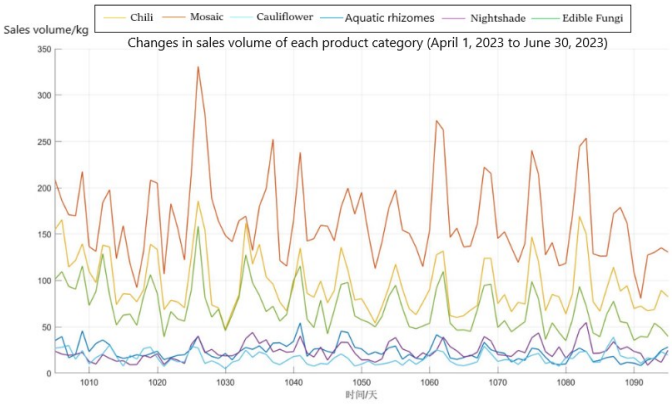

Figure 1: Line chart of sales changes of various types of vegetables in the supermarket from April to June 2023.

We will conduct mathematical modeling based on this data to demonstrate how to analyze the relationship between the total sales volume of each vegetable category and cost-plus pricing, and based on the model, give the daily replenishment volume and pricing strategy for each vegetable

DOI: 10.54254/2755-2721/121/2025.20231

category in the next week (July 1-7, 2023), so as to reduce the waste rate of vegetable category products and maximize the profits of supermarkets.

Before building the mathematical model, we make the following assumptions:

For all vegetable products mentioned in the attachment, if they are not sold on the same day, they will not be sold the next day ; the wholesale price will remain stable and unchanged from July 1, 2023 to July 7, 2023; the transportation costs of various vegetable products are approximately equal; the supply on the supply side is sufficient, and it can be approximately regarded as no limit on the replenishment quantity; the transportation loss rate of vegetable products will remain unchanged in a short period of time; the wholesale price of individual vegetable items will not change in a short period of time.

In order to ensure that the collected data is free of anomalies, we preprocessed the collected data and found no abnormal values or missing values.

2.2. Model establishment

Our goal is to solve two problems. The first is to analyze the relationship between the total sales volume and cost-plus pricing of each vegetable category. The second is to develop a plan for the total replenishment and pricing of each vegetable in the coming week.

For the "cost-plus pricing" method, we consulted the literature and obtained the corresponding pricing formula [4]:

$$\text{Price} = \text{Unit Cost} \cdot (1 + \text{Cost Rate})$$

$$\text{Cost rate} = \text{price/unit cost} - 1$$

2.2.1. Establishment of sales volume-pricing model

The first question is to analyze the relationship between the sales volume and cost-plus pricing of each vegetable category. Based on the basic concept of cost-plus pricing, assuming that the wholesale price is the cost, the following pricing relationship is obtained:

$$\mathrm{p} = \mathrm{c} \cdot (1 + \mathrm{r})$$

p represents the sales price, c represents the unit cost, and r represents the cost interest rate.

By analyzing historical sales data, we hope to establish a quantitative relationship between $i$the average cost rate of each vegetable category $\bar{r}_i(t)$and the total sales volume on each day $m_i(t)$.

The average cost rate $\bar{r}_i$ is calculated as:

$$\bar{r}_i(t) = \frac{P_i(t) - C_i(t)}{C_i(t)} = \frac{P_i(t)}{C_i(t)} - 1$$

Wherein $P_i(t)$、$C_i(t)$they represent $i$the total sales and total cost of the products of the th vegetable category on the day.tAfter data processing, we get $\bar{r}_i(t)$the data table corresponding to the average cost interest rate. the sales volume of various vegetable categories m and the changes in average cost-profit over time from $\bar{r}_i$ on March 27, 2023 (t = 1000) to June 30, 2023 (t = 1095). Take the average cost-profit rate of cauliflower products as an example, which shown in the following figure:

DOI: 10.54254/2755-2721/121/2025.20231

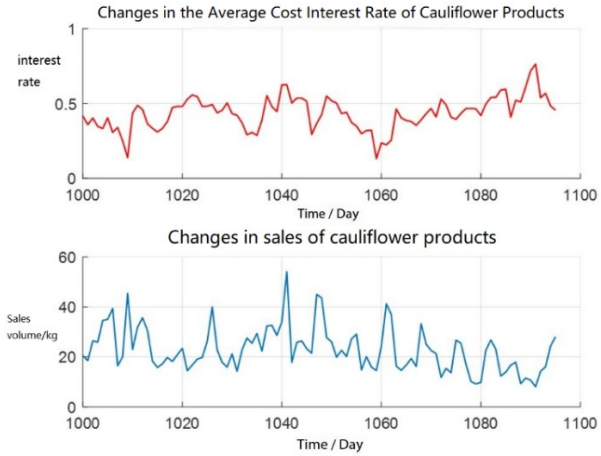

Figure 2: Changes in sales volume and average cost rate of cauliflower products from March 27 to June 30, 2023.

In the following modeling process, $r_i$denoted by

$$[\bar{r}_i(1), \bar{r}_i(2), \bar{r}_i(3), \bar{r}_i(4), \ldots \ldots \bar{r}_i(t), \ldots \ldots]^T,$$

which is $i$the vector corresponding to the average cost rate of the th type of vegetables per day. ($1 \leq i \leq 6$). It represents $[m_i(1), m_i(2), m_i(3), m_i(4), \ldots m_i(t), \ldots]^T$ the vector corresponding to the daily sales volume of the $1 \leq i \leq 6$th type of vegetables.

Use regression analysis to find $m_i$the quantitative relationship between and r_i(1<= i<=6, a total of 6 groups of quantitative relationships). First, we think of using the linear regression model [5] to analyze the relationship between $r_i$and m_i. The mathematical model of linear regression is:

$$m_i = k_i r_i + b_i + \epsilon \ \ (i = 1,2, \ldots 6)$$

Wherein k_i、b_i is the linear regression coefficient and intercept corresponding to the th category i of vegetable products, and $\epsilon$is the random error.

Usually, vegetables are a necessity for many consumers. At the same time, most consumers have an upper limit to their demand for vegetables. Even during the peak period of vegetable demand, the sales volume of vegetable products cannot reach infinity, but should be close to a critical value. The Logistic curve has this characteristic when it changes with the horizontal axis.

Therefore, we also consider using the Logistic curve regression model [6], and the corresponding mathematical model is:

$$m_i = \frac{x_{mi}}{1 + \left(\frac{x_{mi}}{x_{0i}} - 1\right) e^{-h_i r}} + \epsilon$$

(where x_{mi}、x_{0i}、$h_i$is the Logistic curve regression coefficient corresponding to the i-th category of vegetable products)

2.3. Analysis and verification of the sales-pricing model

For the linear regression model used in the above model building, draw its corresponding residual, as shown in the following figure. In order to show our analysis process more conveniently and concisely, we only display the analysis results of cauliflower in the following process.

DOI: 10.54254/2755-2721/121/2025.20231



Figure 3: Linear regression residual analysis of sales volume and cost rate of cauliflower products

From the residual graph, we can see that the residuals are both positive and negative, and the confidence intervals of the residuals mostly include the zero point, which shows that the linear regression is reasonable. There are some red segments in the residual graph (the confidence interval does not include the zero point), and the data corresponding to these segments can be regarded as outliers.

According to the second return value of the regression function of Matlab, we can get the confidence interval (significance level $\alpha = 0.05$) corresponding to the regression coefficient of each regression model.

Table 1: Confidence intervals of linear regression coefficients for each category of vegetables

	Lower bound of confidence interval for regression coefficient$k$	Upper bound of confidence interval for regression coefficient$k$
Cauliflower	12.410	25.126
Mosaic	94.206	153.036
Chili	44.936	66.341
Nightshade	8.398	15.119
Edible Fungi	-30.043	10.348
Aquatic rhizomes	-46.537	-20.887

Based on the above table, the following analysis can be made:

The sales volume of cauliflower, leafy vegetables, peppers, and eggplants is positively correlated with the cost rate, because k the upper and lower bounds of the confidence intervals of their corresponding regression coefficients are greater than 0. This is similar to the economic phenomenon of "increasing demand, insufficient supply, and rising prices" in reality.

The positive and negative relationship between edible fungus vegetable products and cost interest rate is not obvious, because the confidence interval of its corresponding regression coefficient $k$contains the zero point.

The sales volume of aquatic root vegetables is negatively correlated with the cost interest rate, because $k$the upper and lower bounds of the confidence interval of its corresponding regression coefficient are both greater than 0.

DOI: 10.54254/2755-2721/121/2025.20231

2.3.1. Establishment of replenishment-pricing model

Assuming that the past replenishment and pricing strategies were generally reasonable, the supermarket management will tend to make pricing and replenishment decisions that maximize the supermarket's profits. Therefore, the time series forecasting method can be used to predict the optimal pricing size and sales volume of each vegetable category during the period from July 1 to July 7.

Based on historical data, the long short-term memory neural network (LSTM) time series model [7] is used to predict the average cost profit and sales volume $m_i$of six vegetable categories in the seven days from July 1, 2023 to July 7 r_i, 2023. The data results are the optimal replenishment volume and pricing strategy for various vegetable categories. From volume-pricing model, we can get the corresponding average cost rate r_i. Since the daily sales volume of each category of vegetables are not collected, we process the data and obtain the corresponding data table.

2.3.2. Analysis and verification of replenishment-pricing model

We used MATLAB to write the program and used the R MSE graph to evaluate this model. The RMSE graph is the root mean square error, which is one of the indicators used to measure image quality. The smaller the R MSE graph, the more similar the images are. The following is the RMSE graph of sales volume and cost rate of each category of vegetables:

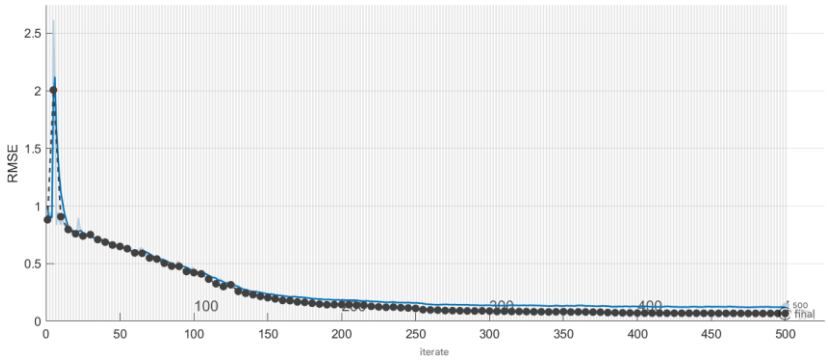

Figure 4: RMSE of cauliflower sales volume

From the above RMSE graph, we can see that the RMSE graphs of sales volume and cost rate of all types of vegetables gradually tend to 0 from the initial large value, reaching a stable state. This shows that the difference between the predicted value of the model and the actual value is very small, and the established model is basically reasonable.

When training the LSTM model, we need to divide the training data into a training set and a test set. We use the training data to train the model and use the test data to select the best model parameters. The following figure is a line graph of the LSTM fitting training data and the actual analysis data, as well as a relative error graph of the training set, drawn using the sales volume and cost rate of each dish as training data:

DOI: 10.54254/2755-2721/121/2025.20231

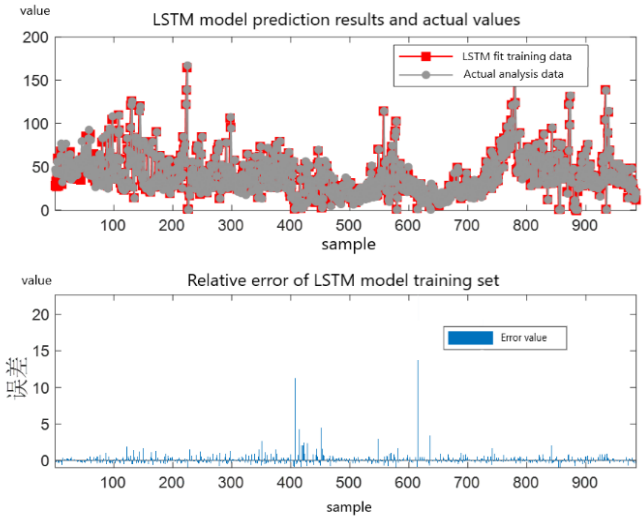

Figure 5: analysis of the training data of cauliflower.

From the above picture, it is known that the line graph drawn by the training set data of the model and the actual value has a high degree of fit, and the relative error between them is mostly only 1-2, which can fully demonstrate that the accuracy of this model is very high.

Next, we can draw a line graph and a relative error graph between the model's test set data and the actual value, as shown in the following figure.

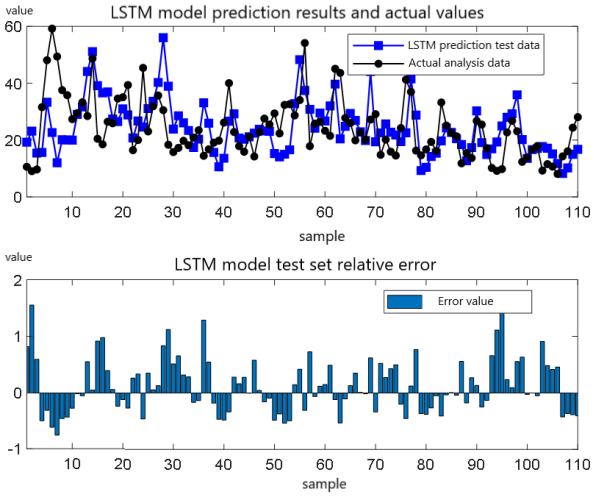

Figure 6: analysis of the test set data of cauliflower.

According to the above figure, the data fitting degree of the test set is relatively high, and the relative error range is mostly between 0 and 1, indicating that the error is small, that is, the model parameters selected by the test set are more accurate, and the prediction accuracy is greatly improved.

3. Results

3.1. Sales-Pricing Model Results

The two regression models mentioned above are solved using the regression function and the lsqcurvefit function of the Matlab software. The solution for the linear regression model is:

DOI: 10.54254/2755-2721/121/2025.20231

Table 2: Linear regression model results

Regression results / Vegetables	b_i (intercept)	k_i (Regression coefficient)	RMSE (Residual Standard Error)
Cauliflower	28.181	18.768	22.553
Mosaic	98.971	123.621	84.990
Chili	48.276	55.639	51.427
Nightshade	13.776	11.758	13.289
Edible Fungi	75.188	-9.848	48.729
Aquatic rhizomes	52.649	-33.712	31.055

The solution of the Logistic curve regression model is:

Table 3: Logistic curve regression equation results.

Regression results / Vegetables	x_{mi}(Maximum)	x_{0i}	h_i(growth rate)	RMSE (Residual Standard Error)
Cauliflower	49.423	16.588	3.822	22.221
Mosaic	183.021	1.185	46.145	85.789
Chili	135.903	35.720	2.471	51.032
Nightshade	21.403	0.044	97.209	12.857
Edible Fungi	70.126	0.000	136.325	48.263
Aquatic rhizomes	37.445	0.073	77.884	31.207

From Table 6 and Table 7, we can see that the residual standard deviation results of the linear regression model are almost the same as those of the Logistic curve model, but the linear regression model needs to fit fewer coefficients and is simpler. Therefore, it can be considered that the linear regression model can better reflect $m_i$the relationship between sales volume and cost interest rate than the r_iLogistic curve regression model.

The linear regression relationship between the sales volume and average cost rate $\bar{r}_i$of various vegetable categories is:

Vegetables: Regression relationship

Cauliflower: $m_1 = 28.181 + 18.768\bar{r}_1$

Mosaic type: $m_2 = 98.971 + 123.621\bar{r}_2$

Chili peppers: $m_3 = 48.276 + 55.639\bar{r}_3$

Nightshade: $m_4 = 13.776 + 11.758\bar{r}_4$

Edible fungi: $m_5 = 75.188 - 9.848\bar{r}_5$

Aquatic rhizomes: $m_6 = 37.445 - 33.712\bar{r}_6$

The corresponding regression coefficient of edible fungi is not significant, which will be mentioned in the following result test.

3.2. Replenishment-Profit Model Results

The sales volume and cost rate for the next 7 days are predicted, and the prediction results are shown in the figure:

DOI: 10.54254/2755-2721/121/2025.20231

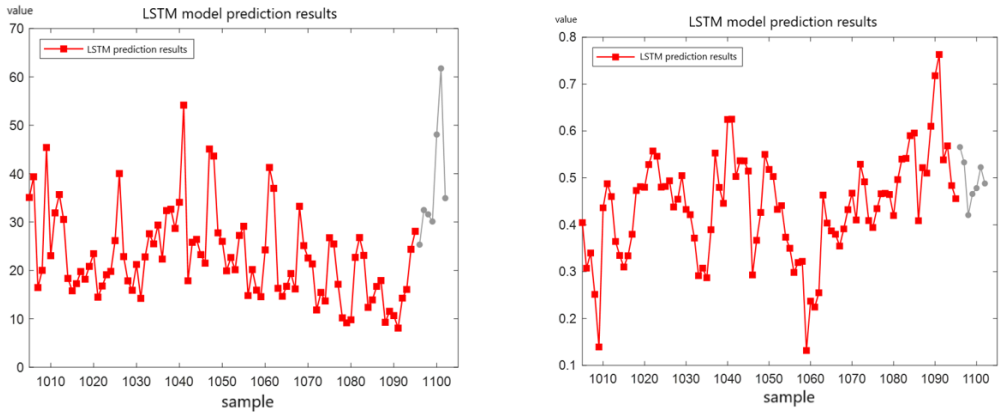

Figure 7: Forecast of sales (left) and cost rate (right) of cauliflower

By analyzing the above prediction result image, we can know the sales volume of each dish in the next 7 days. Since it is known in the question that if the vegetable product is not sold on the same day, it cannot be sold the next day. In order to avoid losses, the daily replenishment volume of the supermarket is equivalent to the sales volume of each dish in the next 7 days, that is, the replenishment volume in the next 7 days is as follows:

Table 4: Daily replenishment of various types of vegetables from July 1 to July 7, 2023.

date	Cauliflower replenish-ment quantity (kg)	Restocking quantity of flowers and leaves (kg)	Chili pepper replenish-ment quantity (kg)	Eggplant replenish-ment quantity (kg)	Edible fungi replenish-ment quantity (kg)	Aquatic rhizomes replenishment quantity (kg)
July 1	25.331	348.288	206.657	23.460	138.377	71.637
July 2	32.487	1171.920	22.175	10.939	129.624	96.794
July 3	31.592	216.580	53.104	17.027	105.721	8.500
July 4	30.123	326.878	145.164	21.415	90.755	20.283
July 5	48.091	316.987	134.934	37.830	88.528	26.101
July 6	61.757	400.437	132.753	31.479	113.264	35.739
July 7	34.941	1109.870	193.720	24.539	172.624	32.466

Similarly, since the selling price of each dish is unknown, we need to predict its cost rate for the next 7 days. In order to obtain the maximum profit, the pricing strategy of the supermarket should give the most appropriate price for each dish every day according to the cost rate, so as to obtain the maximum profit. The results are as follows:

DOI: 10.54254/2755-2721/121/2025.20231

Table 5: Average cost rate forecast for each type of vegetable from July 1 to July 7, 2023

date	Cauliflower	Mosaic	Chili	Nightshade	Edible Fungi	Aquatic rhizomes
July 1	0.566	1.227	0.570	0.700	0.551	0.649
July 2	0.533	1.260	0.519	0.577	0.533	0.534
July 3	0.421	1.212	0.448	0.561	0.435	0.568
July 4	0.466	1.047	0.459	0.576	0.438	0.479
July 5	0.478	1.051	0.571	0.611	0.534	0.531
July 6	0.523	1.209	0.568	0.595	0.583	0.707
July 7	0.488	0.455	0.518	0.544	0.640	0.408

The data in Table 10 is the optimal replenishment strategy for various categories of vegetable products from July 1, 2023 to July 7, 2023 obtained using the prediction model.

The data in Table 11 are the optimal pricing strategies from July 1, 2023 to July 7, 2023 obtained using the forecasting model. For example, the average cost rate data of cauliflower vegetables on July 1, 2023 in Table 11 is 0.566, which means that the supermarket management should control the average profit of cauliflower products at around 56.6 %.

4. Conclusion

This paper proposes an automatic pricing and replenishment decision model that combines statistical analysis, time series prediction and optimization methods, which provides a valuable reference for the sales management of vegetable products. The research results show that the model can not only increase the profits of supermarkets, but also significantly reduce the loss of fresh products, thereby improving the overall supply chain efficiency. Future research can continue to explore the impact of more external factors on the sales of vegetable products to further improve the practicality and adaptability of the model.

The model proposed in this paper has the advantages of being closely integrated with data and concise. Before establishing the model, we simplified the actual situation through relatively reasonable assumptions, making the model easier to solve. In the modeling, we closely followed the collected data for statistical analysis, so that the conclusions have a certain basis. In addition, when analyzing the correlation of sales volume and the relationship between category sales volume and price through the established model, various statistical tests were conducted to make the results more credible.

The model proposed in this article also has room for improvement. The model is based on strong assumptions, which simplify the actual situation. For example, the transportation loss rate remains unchanged, the wholesale price remains unchanged, and only a fixed number of goods are considered, which may deviate from the actual situation. At the same time, we lack more dimensional data, such as transportation time, inventory cost and other factors. If we can further collect multidimensional data, such as data on the significant impact of climate factors such as temperature, humidity, and precipitation on the demand for fresh products. Incorporating these data into the model can improve the prediction accuracy.

References

[1] *Ferguson, M., & Koenigsberg, O. (2007). How should a firm manage deteriorating inventory? Production and Operations Management, 16(3), 306-321.*

[2] *Choi, T.-M., Cheng, TCE, & Wang, SY (2014). Mean–variance analysis of quick response program for managing seasonal products. European Journal of Operational Research, 239(3), 734-744.*

[3] *Kunz, N., Mayers, K., & Van Wassenhove, LN (2019). Closing the loop: Effective sourcing strategies for waste reduction in supply chains. Journal of Cleaner Production, 214, 573-584.*

DOI: 10.54254/2755-2721/121/2025.20231

[4] Han Junhua. Review of cost-plus pricing. Economic Journal. 2012 (22): 74-75.

[5] Li J.Overview of high dimensional linear regression models[C]//Department of Mathematics, Purdue University, Department of Computer and Information Sciences, University of Strathclyde.Proceedings of the 2nd International Conference on Computing Innovation and Applied Physics (part2).Boston University;, 2023:6.DOI:10.26914/c.cnkihy.2023.109800.

[6] Heo, TY., Lee, JM, Woo, MH et al. Logistic regression models for elastic shape of curves based on tangent representations. J. Korean Stat. Soc. 53, 416–434 (2024).

[7] ivieris, IE, Pintelas , E. & Pintelas , P. A CNN–LSTM model for gold price time-series forecasting. Neural Comput & Applic 32, 17351–17360 (2020).