# Pricing and replenishment decision problem based on time series prediction and optimization algorithm

Zijing Li*
School of mathematics and statistics,Wuhan University, Wuhan, China, 430072
* Corresponding Author Email: 2021302142022@whu.edu.cn

**Abstract.** The pricing of vegetable products is an important research topic that involves multiple aspects such as agricultural production, market supply and demand, consumer interests, and social benefits. Vegetable products have characteristics such as short production cycles, short shelf life, large price fluctuations, and high demand elasticity, which pose challenges to pricing decisions. To develop replenishment decisions that maximize the profits of supermarkets, the author establishes a mathematical model for the total sales volume and cost add-on pricing of each vegetable category. First, the cost, cost add-on coefficient, and loss rate of individual products in the same category are summed according to the weight of sales, and the corresponding average cost, cost add-on coefficient, and average loss rate of the category are obtained. This initially achieves a negative correlation between the sales trend and the cost add-on coefficient using a function. Then, by constructing a LSTM neural network model and using known data to predict the sales of each category, the predicted sales volume of the model is used as the sales benchmark value. The relationship function between the sales trend and the cost add-on coefficient is supplemented, and the optimization algorithm is used to find the cost add-on coefficient value, the total daily replenishment quantity, and the maximum profit value that maximize the profits of supermarkets.

**Keywords:** supermarket replenishment, cost add-on coefficient, LSTM model.

## 1. Introduction:

### 1.1. Research Background and Significance

In the study of vegetable pricing and replenishment decision-making, researchers have obtained conclusions on the impact of factors such as traditional retailer prices and inventory levels on vegetable sales. However, research on replenishment and preservation conditions is slightly lacking and a good theoretical expression has not been sought. Vegetables, as well as daily consumables, are indispensable and indirectly impact the socio-economy. Therefore, the author analyzes the regularities and relationships of vegetable categories and individual products, comprehensively applies forecasting and optimization models, and establishes replenishment and pricing decision-making models with the objective of maximizing supermarket revenue. Based on market rules, the author suggests that supermarkets need to further collect data to promote the practical application of new models constructed from these data.

### 1.2. Literature Review

In the article “Joint Decision on Pricing and Inventory Replenishment of Agri-food with Dual-channel Sales” the researchers establish a joint decision model of pricing and inventory with demand dependent on price and inventory level and draw the conclusion that the retailer profits increase with the extension of the fresh-keeping period, and fluctuate slightly with the increase of the deterioration cost under dual-channel sales. The basic conditions for retailers to gain higher profits are lower fresh-keeping cost and deterioration cost, higher sales prices and a larger online market share. The author also finds out that with reduced vegetable spoilage rate, supermarkets can gain more profits. The author and the article’s conclusion share something similar[1].

# 2. The relationship between category sales volume and cost add-on pricing and the profit maximization model

## 2.1. Research Content and Framework

The author collects data of 5 kinds of vegetables: eggplant,aquatic root and tuber,edible mushroom,chili,cauliflower and leafy[2].To solve the problem, the author divides the problem into two parts for model analysis, establishment and solution.

Part 1: Based on the category data, the author analyzes the relationship between the sales volume and cost add-on pricing of vegetable categories by category.

Part 2: With the maximum profit of the supermarket as the objective function, the author determines the daily replenishment volume and pricing strategy for July 1-7, 2023.

**Part 1: The model of the relationship between vegetable category sales volume and cost-plus pricing**

**Method:**For part 1, the author gives the definition and calculation method of "cost add-on coefficient" and "sales volume of each vegetable category". Then calculates the values of these two quantities. Then draws the image of the sales volume and cost add-on coefficient of each vegetable category. Finally, the author analyzes the relationship between them according to the image, draws conclusions and makes result analysis.

The author determines that the method of "cost add-on pricing" can be expressed as:

Product price = cost × cost add-on coefficient,

where, the "cost add-on coefficient" is denoted by $r$.

For each item, the cost add-on coefficient of item $i$ is the ratio of the price and wholesale price of that item, that is:

$$r_i = \frac{b_i}{a_i}, \tag{1}$$

where $a_i, b_i$ are the wholesale price and price of item $i$, $y_i$is the sales volume of item $i$. For categories, since a category contains many items, to characterize the impact of each item on the overall category, the cost add-on coefficient of the category is defined as the weighted average of $r_i$ for each item in that category. Let $\bar{r}$ be the weighted average of the cost-plus pricing $r$ for that category, then its expression is:

$$\bar{r} = \frac{\sum r_i y_i}{\sum y_i}, \tag{2}$$

where $y_i$ is the sales volume of item $i$.

Add up the sales volume of all vegetable items belonging to a certain category in a day, and get the sales volume of a certain vegetable category $Y$. According to the data, draw the graph of the sales volume and cost add-on coefficient of each vegetable category and analyze the relationship between them based on the graph.

**Result:** This article shows the relationship between the sales volume and cost add-on coefficient of cauliflower and eggplant categories as shown in Figure 1. The graphs of other categories are similar.

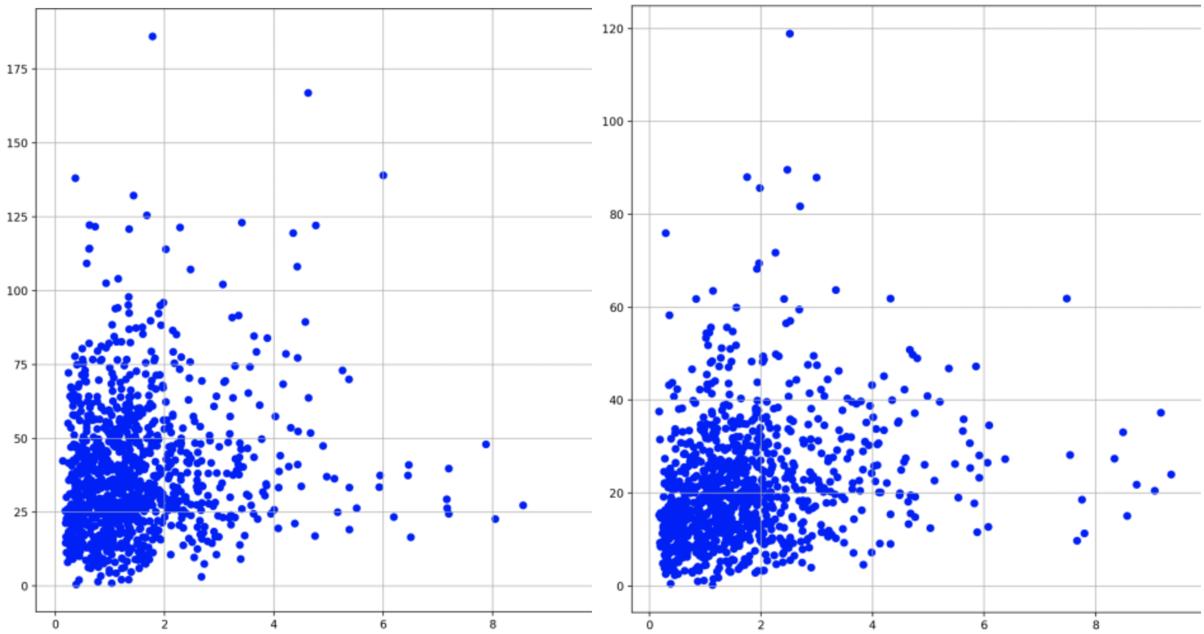


**Figure 1.** Scatter plot of sales and cost add-on coefficients for Cauliflower and Eggplant vegetables

**Discussion**:The data points are densely distributed in the lower left corner, that is, most of the data points have small horizontal and vertical coordinates. Combined with the actual meaning, the category sales volume is mostly below 150kg, and the cost add-on coefficient mostly does not exceed 5, which is consistent with the market pricing rules and sales volume data that the author consults.

Observing the outliers, fitting these points into a curve can see that it shows a negative exponential decay relationship, that is, when the cost add-on coefficient increases, the sales volume will tend to decrease, which is consistent with the market rules. In order to find out a more accurate relationship between vegetable categories and cost add-on pricing, the author first analyzes the scatter plot as follows:

In the densely populated lower left corner area, a small area may overlap many different time and sales volume information, so only by observing the lower left corner area can not draw a suitable conclusion. The author needs to combine the external points in the dense area and the overall rules to analyze the relationship between sales volume and cost add-on pricing.

It is observed that when the sales volume increases, that is, when the horizontal coordinate extends to the right, the cost add-on coefficient gradually decreases, and the outliers of the vertical coordinate gradually approach the x-axis. When the sales volume is moderate, as the sales volume increases, the value of the cost add-on coefficient decreases faster, while when the sales volume is large (exceeds 150kg), as the sales volume increases, although the value of the cost add-on coefficient decreases, but the overall area is stable. Fitting curves to peripheral points shows that sales volume and cost add-on coefficient are close to negative exponential relationships[3].

**Part 2: The supermarket profit maximization decision model**

**Method:** For part 2, combining the prediction model and the optimization model. First, the author selects a more suitable LSTM model for predicting the sales volume baseline value according to the data characteristics. Second, the author establishes the optimization model, and determines the two decision variables as the cost add-on coefficient $r$ and the sales volume $Y$, and determines the objective function as the maximum profit of the supermarket. Then the author uses the relationship between the variables to give the expression of the objective function and other constraints. Finally, solves the optimization problem and draws conclusions, and gives result analysis[4].

To test the rationality of the model, the author predicts the data for the next week based on the past data. The author uses the LSTM model, which can better solve the long-term dependency problem and avoid the gradient vanishing or exploding problem in the traditional prediction model, to predict the sales volume, and takes the predicted sales volume as the baseline value[5].

Let $y$ be the predicted sales baseline value. Let $Y$ be the sales volume, which can be considered to satisfy the following formula:

Sales volume $Y$ = Predicted sales benchmark value $y$ × Sales coefficient $k$,

where the sales coefficient is a function of the cost add-on coefficient, and this sales formula can reflect the impact of the change in the cost add-on coefficient on sales. Then the author determines the relationship between the sales coefficient $k$ and the cost add-on coefficient $r$ for a specific category, assuming that $k$ and $r$ satisfy some relationship:

$$k = g(r), \tag{3}$$

then $g(r)$ should satisfy the following conditions:

$$\begin{cases} g'(x) < 0 \\ g(1) = +\infty \\ g(+\infty) = 0 \\ g(\bar{r}) = 1 \end{cases} \tag{4}$$

By observing the curve graph of sales volume and cost add-on pricing of single product, it is preliminarily determined that $r$ has a negative exponential relationship with $g(r)$. The author concludes that the expression of $g(r)$ is:

$$g(r) = \exp\{-\frac{r-\bar{r}}{(r-1)^l}\} \tag{5}$$

By changing the value of $l$ in the expression continuously and comparing it with a large number of single product curve graphs, it is finally found that the fitting effect is better when $l = 0.2$. Therefore, the author determines the expression of the sales coefficient $g(r)$ as:

$$g(r) = \exp\{-\frac{r-\bar{r}}{(r-1)^{0.2}}\} \tag{6}$$

The function relationship of $g(r)$ at different $\bar{r}$ values is shown in Figure 2.

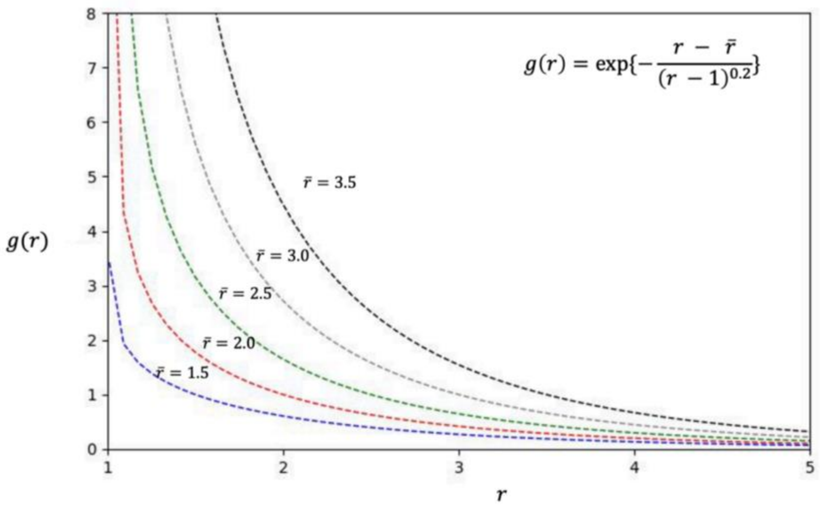


**Figure 2.** Line graph depicting the variation of sales coefficients with cost add-on coefficients

Taking each category as a unit, regard each category as a whole, and calculate the average cost $c$ of the category, so that the average profit of the category is $rc$. The loss rate is also taken as the average value $\delta$ and the expression of $\bar{\delta}$ is as follows:

$$\bar{\delta} = \frac{\sum \delta_i y_i}{\sum y_i} \tag{7}$$

By the assumption, the sales volume is:

$$Y = yg(r) \tag{8}$$

The total replenishment quantity is:

$$X = \frac{yg(r)}{1-\bar{\delta}} \tag{9}$$

Therefore, the profit $W$ is obtained by subtracting the cost from the sales amount:

$$W = \exp\{-\frac{r-\bar{r}}{(r-1)^{0.2}}\}(r - \frac{1}{1-\bar{\delta}})yc \tag{10}$$

The author determines the objective function of the optimization model as:

$$\max W$$

$$W = \exp\{-\frac{r-\bar{r}}{(r-1)^{0.2}}\}(r - \frac{1}{1-\bar{\delta}})yc \tag{11}$$

So the pricing method is as follows: This question takes the category as the unit, so it can be assumed that the difference between the products within the same category is small. According to the calculated cost-plus coefficient for each category, the same cost-plus coefficient is used to determine the price for all products within that category. In addition, in the evening, the cost-plus coefficient should also be adjusted appropriately according to the remaining quantity, appearance, quality, etc. of the product, and a discount strategy should be adopted.

Then the author applies LSTM model in forecasting and attains the results.

To predict the data for 7 days, the LSTM (Long Short-term memory) model is chosen to forecast the sales baseline value, because the LSTM model can retain the contextual information in the sequence through its neural network "internal loop"[6], and has the ability to retain long-term memory and avoid gradient disappearance[7]. The implementation steps of the LSTM model are as follows[8]:

**Step 1:** Decide what information to discard from the cell state. This decision is made by a sigmoid layer called the "forget gate layer". It receives $h_t$ and $x_t$, and outputs a value between 0 and 1 for each number in the cell state $C_{t-1}$. 1 means "keep it completely", 0 means "discard it completely".

**Step 2:** Calculate the input gate $i_t$, which is used to control what information to add to the cell state. The formula for $i_t$ is:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{12}$$

where $\sigma$ is the sigmoid function, $W_i$ and $b_i$ are the parameters of the input gate.

**Step 3:** Calculate the forget gate $f_t$, which is used to control what information to delete from the cell state. The formula for $f_t$ is:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{13}$$

where $W_f$ and $b_f$ are the parameters of the forget gate.

**Step 4:** Calculate the candidate cell state $C_t$, which is used to generate new information. The formula for $C_t$ is:

$$C_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_C), \tag{14}$$

where $\tanh$ is the hyperbolic tangent function, $W_C$ and $b_C$ are the parameters of the candidate cell state.

**Step 5:** Update the cell state $c_t$, which is used to store long-term memory.

**Step 6:** Calculate the output gate $o_t$, which is used to control what information to output to the hidden state. The formula for $o_t$ is:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{15}$$

where $W_o$ and $b_o$ are the parameters of the output gate.

**Step 7:** Calculate the hidden state $h_t$, which is used to store short-term memory.

**Result:** The maximum sales profit for each category from July 1 to 7, 2023 is shown in the Table 1.

**Table 1.** Maximum sales profit for each category from July 1st to July 7th, 2023

| | 2023.7.1 | 2023.7.2 | 2023.7.3 | 2023.7.4 | 2023.7.5 | 2023.7.6 | 2023.7.7 |
|---|---|---|---|---|---|---|---|
| Leafy | 2046.42 | 2075.25 | 2110.11 | 2148.35 | 2187.92 | 2227.82 | 2267.27 |
| Cauliflower | 105.99 | 103.58 | 101.59 | 99.75 | 97.86 | 95.93 | 94.00 |
| Aquatic root and tuber | 77.88 | 73.27 | 66.78 | 60.31 | 53.97 | 47.78 | 41.71 |
| Eggplant | 102.97 | 104.47 | 105.36 | 105.91 | 106.30 | 106.60 | 106.85 |
| Chili vegetables | 482.05 | 479.74 | 474.50 | 467.59 | 460.25 | 452.94 | 445.78 |
| Edible mushroom | 338.58 | 336.63 | 337.26 | 337.63 | 337.32 | 336.60 | 335.64 |

The daily replenishment volume corresponding to the maximum profit of each category from July 1 to 7, 2023 is shown in Table 2.

**Table 2.** Daily replenishment quantity corresponding to the maximum sales profit for each category from July 1st to July 7th, 2023

| | 2023.7.1 | 2023.7.2 | 2023.7.3 | 2023.7.4 | 2023.7.5 | 2023.7.6 | 2023.7.7 |
|---|---|---|---|---|---|---|---|
| Leafy | 1066.73 | 1081.76 | 1099.93 | 1119.87 | 1140.49 | 1161.29 | 1181.85 |
| Cauliflower | 16.22 | 15.85 | 15.54 | 15.26 | 14.97 | 14.68 | 14.38 |
| Aquatic root and tuber | 10.86 | 10.21 | 9.31 | 8.41 | 7.52 | 6.66 | 5.82 |
| Eggplant | 17.43 | 17.69 | 17.84 | 17.93 | 17.99 | 18.05 | 18.09 |
| Chili vegetables | 93.33 | 92.88 | 91.87 | 90.53 | 89.11 | 87.69 | 86.31 |
| Edible mushroom | 75.58 | 75.14 | 75.28 | 75.36 | 75.30 | 75.14 | 74.92 |

For each category, the $r$ value taken every day remains unchanged. The $r$ values corresponding to the 6 categories are shown in Table 3.

**Table 3.** Cost add-on coefficient corresponding to the maximum sales profit for each category from July 1st to July 7th, 2023

| | Leafy | Cauliflower | Aquatic root and tuber | Eggplant | Chili vegetables | Edible mushroom |
|---|---|---|---|---|---|---|
| Cost add-on coefficient | 1.79 | 2.3 | 2.27 | 2.19 | 2.1 | 1.96 |

After obtaining the solution results, the author tests the prediction effect. The fitting effect graph of the sales volume of cauliflower category on the last test set is shown in Figure 3. As can be seen from the graph, the prediction value is close to the true value, indicating that the model has a relatively good fitting effect.

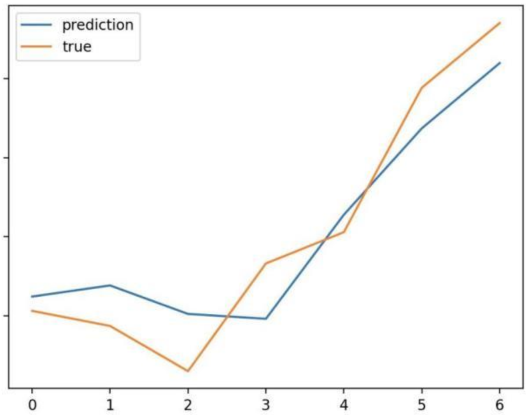


**Figure 3.** Fitting effect graph of cauliflower sales volume

The optimization process of the RMSE value of the LSTM model is shown in Figure 4. As can be seen from the figure, as the model is optimized and improved, the RMSE gradually decreases, and the final RMSE value is 0.0270, indicating that this model has a good fitting effect.

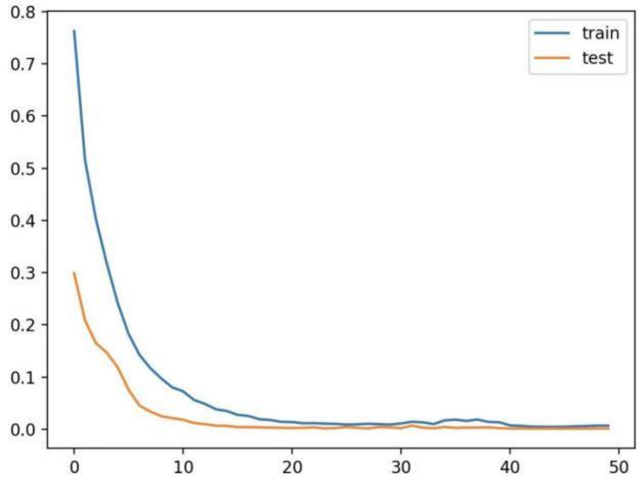


**Figure 4.** Optimization process of LSTM model's RMSE value

**Discussion:** For the same category, the calculated cost add-on coefficients are all the same value. In fact, finding the maximum point of the objective function $W$ is equivalent to finding the maximum point of $\exp\{-\frac{r-\bar{r}}{(r-1)^{0.2}}\}(r-\frac{1}{1-\bar{\delta}})yc$, which only depends on the category of the category, and is independent of the number of days and the predicted sales base value. Therefore, the same value of the cost add-on coefficient will be obtained.

The RMSE value shows that this model has a good prediction effect. The reasons for the more accurate prediction are[9]:

a. The LSTM model can capture the long-term dependency relationship in the data better by using cell state and gate mechanism, has a better memory function, and can retain more distant context information when processing sequence data;

b. The LSTM model overcomes the problems of gradient explosion and gradient disappearance that often occur in traditional models, ensuring its accuracy when processing longer sequence data.

The author believes that by introducing a multi-party competition model, it is possible to determine a pricing range that conforms to the principles of free market competition. Based on the relationship between total sales volume and the determined pricing, the total replenishment quantity can be determined. This pricing model allows for more reasonable pricing in supermarkets and ensures that the replenishment quantity is moderate, thereby avoiding resource wastage or the market drawbacks of severe supply-demand imbalance.

## 3. Conclusion

For the same category, the calculated cost add-on factors are all the same. In fact, maximizing the objective function W is equivalent to maximizing $\exp\{-\frac{r-\bar{r}}{(r-1)^{0.2}}\}(r-\frac{1}{1-\bar{\delta}})$, which is only dependent on the category and independent of the number of days and the forecast sales benchmark. Therefore, it will yield the same value for the cost add-on factor[10]. From the RMSE value, it can be seen that the model has good prediction performance. The reasons for the accurate predictions can be analyzed as: the LSTM model, through its cell state and gate mechanisms, is able to effectively capture long-term dependencies in data, providing excellent memory capabilities and preserving distant contextual information when processing sequential data. Meanwhile, the LSTM model overcomes the problems of gradient explosion and vanishing gradients that often occur in traditional models, ensuring its accuracy when dealing with longer sequence data.

## References

[1] Chen Jun, Kang Sha. Joint Decision on Pricing and Inventory Replenishment of Agri-food with Dual-channel Sales[J]. Industrial Engineering Journal, 2023, 26(3): 39-46.

[2] Zhihu User. Saved: Classification, season, cold and hot attributes, and selection tips for vegetables. Zhihu, 2021-03-23. https://zhuanlan.zhihu.com/p/359343261.

[3] Wang Haopeng. Research on Inventory Forecasting and Replenishment Strategy Optimization Method for Pharmaceutical Logistics Warehousing[J], Northwest Normal University, 2023.

[4] Chen Zewei. Research on Inventory Replenishment Strategy of Agricultural Product Retailers Considering Different Behaviors[J],Chongqing Jiaotong University, 2023.

[5] Zhao Han. Research and application of multi-step prediction method for time series data based on RNN [D]. Beijing University of Technology, 2020. DOI: 10.26935/d.cnki.gbjgu.2020.000251.

[6] Yu Y, Si X, Hu C, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. Neural computation, 2019, 31(7): 1235-1270.

[7] Staudemeyer R C, Morris E R. Understanding LSTM–a tutorial into long short-term memory recurrent neural networks[J]. arXiv preprint arXiv:1909.09586, 2019.

[8] Zhihu User. LSTM that everyone can understand.2018. https://zhuanlan.zhihu.com/p/32085405

[9] Jiang Qiyuan, Xie Jinxing, Ye, Jun. Mathematical Modeling [M]. 4th edition, Beijing: Higher Education Press.

[10] Huang Lixue, Yao Guanxin. Research on horizontal supply of fresh agricultural products in chain supermarket[J]. Logistics Sci-Tech, January,2024(the first half).