Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 6643195, 19 pages
https://doi.org/10.1155/2021/6643195

Research Article

Joint Strategy of Dynamic Ordering and Pricing for Competing Perishables with Q-Learning Algorithm

Jiangbo Zheng,[1] Yanhong Gan,[2] Ying Liang,[3] Qingqing Jiang,[1] and Jiatai Chang[1]

[1]*School of Management, Jinan University, Guangzhou, 510632 Guangdong, China*
[2]*School of Business Administration, South China University of Technology, 510640 Guangzhou, China*
[3]*School of Economics and Management, South China Normal University, Guangzhou, 510006 Guangdong, China*

Correspondence should be addressed to Ying Liang; 20131196@m.scnu.edu.cn

Received 13 December 2020; Revised 28 January 2021; Accepted 14 February 2021; Published 13 March 2021

Academic Editor: Wenqing Wu

We use Machine Learning (ML) to study firms' joint pricing and ordering decisions for perishables in a dynamic loop. The research assumption is as follows: at the beginning of each period, the retailer prices both the new and old products and determines how many new products to order, while at the end of each period, the retailer decides how much remaining inventory should be carried over to the next period. The objective is to determine a joint pricing, ordering, and disposal strategy to maximize the total expected discounted profit. We establish a decision model based on Markov processes and use the Q-learning algorithm to obtain a near-optimal policy. From numerical analysis, we find that (i) the optimal number of old products carried over to the next period depends on the upper quantitative bound for old inventory; (ii) the optimal prices for new products are positively related to potential demand but negatively related to the decay rate, while the optimal prices for old products have a positive relationship with both; and (iii) ordering decisions are unrelated to the quantity of old products. When the decay rate is low or the variable ordering cost is high, the optimal orders exhibit a trapezoidal decline as the quantity of new products increases.

1. Introduction

Due to the scarcity of resources and the advance of technology, both academia and practice have highlighted the significance of value deterioration, focusing on perishables as a central issue. Many previous studies on perishable products focused on food items (e.g., meat, poultry, produce, dairy, and bakery products), pharmaceuticals (e.g., drugs and vitamins), and cut flowers; this paper studies perishables caused by the replacement of high-tech products, such as laptop computers and telephones, have an increasingly short life cycle due to rapid advances in science and technology and thus exhibit perishable characteristics. It is more convincing to study the joint dynamic ordering and pricing of high-tech products in multiperiods than traditional static pricing. According to a survey conducted by *China Youth Daily* in April 2018, 71.8% of respondents changed their mobile phones at least once every two years and 42% said that their still in use digital products should be updated in a higher frequency.

Managing perishables remains a significant but open issue for industry and academia. A key feature of perishable inventory systems is that each product has its own finite shelf-life; retailers may even have an inventory of a single product with a wide variety of ages on the same shelf. Therefore, the biggest challenge may stem from matching the supply of variously aged perishable goods with the diversity of customers' demand. For instance, some customers like high-quality products while others prefer low prices. Retailers could thus do more to ensure that the quantity of new products meet the demands of consumers with quality sensitivity while discounting older products to attract consumers with price sensitivity when considering the dynamic demand competition between different ages' products, with the purpose of maximizing their profits. For example, when Apple introduces new mobile phones, it will continue to sell old phones at low prices; when the car launches a new model, the old model will continue to sell at a reduced price. How should these retailers dynamically develop joint ordering and pricing strategies when considering the purchase

behavior of heterogeneous consumers and dynamic demand substitution between different ages' products in a multiperiod? However, many studies have neglected the impact of the number of old products carried into the next period on joint strategy with assuming that all remaining valuable old products are entered to the next period or all are discarded. This paper will verify the optimal dynamic data of old products carried into next period as well as the optimal dynamic ordering and pricing strategy in a multiperiod when different ages' products sale simultaneously. Further, this paper is aimed at answering the complicated questions as follows. How should retailers make a trade-off between decreasing order quantity to reduce ordering costs and increasing quantity to accrue a greater share of profit? How should retailers make a trade-off between discarding valuable old products and swallowing new products' profits?

Specifically, we consider two different types of products, new and old, each with its own set of qualities, being sold simultaneously in a given period. The retailer must vary the quantity and price of the two products to maximize long-term revenue when considering dynamic demand competition, positive lead time, fixed ordering cost, and inventory holding cost. In addition, we introduce a consumer utility function to develop a demand model that considers customer purchase behaviors and build our Markov decision model to obtain an optimal strategy, including the decision actions selected for each state that can maximize the total expected discounted profit over an infinite horizon. Then, we originally develop a reinforcement learning algorithm, Q learning, which can solve optimal strategies more effectively without the state transition probability. Finally, we analyze changes to the optimal strategy under different parameters and then present management implications corresponding to our numerical experiments. We find that by incorporating the number of old products carried into the next period in the decision-making, retailers have greater flexibility and more decision options. Although each strategy presents complex features, the Markov decision-making model has a more comprehensive reference value.

The remainder of the article is structured as follows. In Section 2, we review the related literature. Section 3 describes the cost and demand models (the latter based on the vertical differentiation model in Moorthy, 1988) and develops the final model based on Markov decision processes. In Section 4, we use the Q-learning algorithm to analyze the model's optimal strategy; the characteristics of which are illustrated through numerical analyses. Finally, Section 5 presents managerial implications and offers directions for future research.

2. Literature Review

The literature on ordering and pricing strategies for perishable inventories is substantial, while very few studies combine ordering, pricing, and disposal policies in a single model validating different ages' products sold in the same period. Hence, the literature streams most relevant to our research concerns involve (i) inventory control for perishables, (ii) dynamic pricing for perishable inventories, and (iii) customer behavior modeling for ordering, pricing, and other controls.

The research on inventory control for perishables involves two types of perishability: random and fixed lifetime. Fixed lifetime models were pioneered by Nahmias and Pierskalla, who considered inventory control for perishables with a two-period lifetime and discussed the optimal ordering strategy within a finite period without a detailed solution [1]. When a product's lifetime is longer than two periods, ordering strategies involving old products have complex, nonlinear structures. This approach was developed by Fries and Nahmias, but both of these studies involve trade-offs: lost sales in the former and delayed deliveries in the latter [2, 3]. Researchers therefore sought more effective heuristic strategies to solve these problems. Nandakumar and Morton studied a heuristic ordering strategy for multilifetime perishables based on a discrete time-inventory model. He proposed and verified a heuristic strategy for retailers to determine the upper and lower bounds of order quantity [4]. Li et al. studied joint replenishment and clearance sales for perishables under a general finite lifetime and a last-in-first-out (LIFO) issuing rule. They found that optimal strategies can be characterized by two product-inventory thresholds and proposed that products with different lifetime remainders be sold at clearance sales [5]. When considering fixed ordering costs and lead time constraints, inventory strategies are likely to have to be adaptable, as retailers face a trade-off between the frequency and quantity of orders based on cost. Coelho and Laporte introduced the joint replenishment and inventory control of perishable product, while considering inventory holding costs, disposal costs, they modeled the problem under general assumptions as a MILP and solved it exactly by branch and cut [6]. Berk and Gürler used the Markov decision process to develop their model and evaluated it using the (Q, r) strategy. In this way, they were able to study ordering strategies' various parameters for multilifetime perishable products with positive lead time, allowing for lost sales [7]. Chao et al. developed an approximation algorithm for perishable inventory systems with positive lead times and finite ordering capacities, showing that their model admits a theoretical worst-case performance under positively correlated demand processes [8]. Our study builds on this work to focus on perishable inventory systems with a fixed lifetime, positive lead time, and fixed ordering costs. However, it is insufficient to study only the inventory control problem for perishables. Ordering and pricing strategies influence each other, and it is necessary to incorporate both into the decision-making process to maximize the retailer's revenue.

There is a rapidly growing stream of literature on dynamic pricing and inventory control for perishables. Feng studied an optimal replenishment model with dynamic pricing and quality investment for perishable products, and the dynamic optimization model is proposed to maximize the total profit per unit time and solved on the basis of Pontryagin's maximum principle [9]. Kaya and Polat investigated the problem of jointly determining the optimal pricing and inventory replenishment strategy for a deterministic perishable inventory system, in which demand is both time and

price dependent. Their model also determined the optimal point to adjust prices in relation to the optimal price and the optimal order quantity [10]. Rabbani et al. discussed optimal dynamic pricing and replenishment policies for items that exhibit deterioration in both quality and physical quantity, where the selling price was defined as a time-dependent function of the initial price and discount rate [11]. When the decision cycle involves more than two periods, the joint pricing and inventory control solutions become extremely difficult. Only a few studies limitedly address this issue. Li et al. considered a dynamic joint pricing and inventory control problem for a perishable product over an infinite horizon, assuming a linear price-response demand model with backlogging and zero lead time. They characterized the optimal strategy's structure as having a two-period lifetime and developed a base-stock/list-price heuristic for stationary systems with multiperiod lifetimes [12]. Li et al. proposed a stationary structural policy consisting of an inventory order-up-to level, state-dependent price, and inventory-clearing decisions and developed a fractional programming algorithm to compute the optimal policy in an infinite-horizon lost-sales case, in which the retailer does not sell new and old inventory at the same time [13]. Chen et al. analyzed joint pricing and inventory control problem for a perishable product with a fixed lifetime over a finite period when different ages' perishables sale simultaneously, but they uniformly priced them [14]. All these papers still have not developed model validation of consumer behaviors and the competition between new and old products.

Finally, our study relates to research that models customer behaviors in the context of ordering, pricing, and other controls. The consumer utility function in Smith et al. provides a reference for dealing with vertical differentiation between different ages' perishables [15]. Ferguson and Koenigsberg studied a two-period joint pricing and inventory control problem, addressing the impact of competition between new and old inventory in a secondary period [16]. Akçay et al. considered a dynamic pricing problem over a finite selling horizon, in which the firm has an initial inventory of multiple substitutable and perishable products. They modeled the multiproduct dynamic pricing problem as a stochastic dynamic program, analyzing its optimal prices with an integrative model of consumer choice that was based on linear random consumer utilities [17]. Sainathan cited consumer utility functions to describe the utility evaluations among different ages' products. He determined the retailer's optimal procurement and pricing strategies for different ages' perishable products over an infinite horizon using a Markov decision problem [18]. Chew et al. determined order quantities and prices for a perishable with a multiple period lifetime, allowing substitution between different ages' products. Specifically, they modeled demand for different ages' products as dependent on both their own prices and the prices of substitutable products, i.e., products of "neighboring ages." They used a stochastic dynamic programming model to obtain the optimal joint policy for a two-period lifetime; however, they failed to provide a specific method for products that have lifetimes longer than two periods [19].

In summary, although there are many studies on perishable inventories, there is a research gap that addresses both dynamic demand substitution and joint ordering, pricing, and disposal strategy for different ages' products in a multiperiod, also lacking of precise algorithms to gain the optimal strategy. On this basis, our contribution to bridge the existing research gap involves the following: (i) we incorporate the number of old products carried to the next period into the joint strategy to better cope with consumer preferences and dynamic demand substitution, with the purpose of maximizing the retailers' profits when considering fixed order cost and inventory holding cost, which is not included in Sainathan [18]; (ii) we develop the Q-learning algorithm rather than dynamic programming or value iteration to solve the Markov model and gain the multiperiod optimal strategy. This algorithm can obtain a stable optimal policy effectively, including the actions of all states, after being trained without the state-transition probability and current expected return. Some results from multiple numerical experiments can provide theoretical guidance to retailers' daily decisions.

3. Model

We consider a retailer who sells a perishable product with a two-period lifetime over an infinite horizon. The retailer has a chance to order in each period, and the lead time is one period. At the beginning of each period, the retailer observes the initial inventory, which includes both old products from the previous period and new products, and decides how many new products to order for the next period. The total number of old and new products on the shelves at any given time cannot exceed the retailer's capacity. The retailer then decides the prices for the new and old products. There are N customers in each period, whose arrival follows a Poisson distribution. Each customer purchases up to one product to maximize utility based on price and quality sensitivity, which follows 0-1 uniform distribution. Because the two products are vertically differentiated, the retailer loses more from trying to sell old products due to their lower price. Hence, the retailer faces a trade-off between product spoilage and demand substitution and thus must take into account the optimal number of old products carried into the next period. The objective of the retailer is to maximize the total expected discounted profit over an infinite horizon. The notations and decisions defining the model are provided in the next sections.

3.1. Notations. λ: average number of arrivals,

N: total number of arrivals in each period following a Poisson distribution,

I: the capacity on shelf,

c_f: fixed ordering cost,

c_v: ordering cost per unit,

c_I: inventory holding cost per unit,

τ_n: the new product quality,

τ_o: the old product quality,

θ: customer's quality sensitivity following 0-1 uniform distribution,

o_r^t: the number of old products remaining at the beginning of period t,

n^t: the number of new products available at the beginning of period t,

g^t: $g^t \in \{0, 1\}$, $g^t = 1$ if the retailer ordered in period t, otherwise $g^t = 0$,

U_{mn}^t: the m customer's utility for new products in period t,

U_{mo}^t: the m customer's utility for old products in period t,

x_{mn}^t: the m customer's purchase of new product in period t,

x_{mo}^t: the m customer's purchase of old product in period t,

d_n^t: the demand of new product in period t,

d_o^t: the demand of old product in period t,

C^t: total cost in period t,

R^t: net profits in period t.

3.2. Decisions. q^t: the order quantity in period t,

o_s^t: the number of products unsold in period t and carried over to period $t + 1$,

p_n^t: the price of new product in period t,

p_o^t: the price of old product in period t.

3.3. Demand Model. Any customer who visits the retailer has three choices—buy one old product, buy one new product, or buy nothing. All customers make their decisions based on utility and preference. Similar to Sainathan utility function [18], let $U_{mn}^t = \theta_m^t \tau_n - p_n^t$ be the utility of a new product for customer m, and $U_{mo}^t = \theta_m^t \tau_o - p_o^t$ be the utility of the old product.

If it meets two conditions, customer m will choose product i:

$$\begin{cases} U_{mi}^t = \max\limits_{j=n,o} U_{mj}^t, \\ U_{mi}^t > 0. \end{cases} \tag{1}$$

According to the above two conditions and granted that the old product will be priced lower than the new product, we can obtain the upper and lower bounds for each product's price:

$$0 < p_o^t < \min\left(\tau_o, p_n^t\right), 0 < p_n^t < \tau_n. \tag{2}$$

There are six situations, and we discuss them independently.

(1) When $U_{mn}^t > 0 \geq U_{mo}^t$ or $U_{mo}^t > 0 \geq U_{mn}^t$, customer m will only purchase new products (or old products) and buy nothing if they are out of stock

(2) When $U_{mn}^t \geq U_{mo}^t > 0$, customers prefer to purchase new products and buy old products if the new products are out of stock; otherwise, they buy nothing. When $U_{mo}^t > U_{mn}^t > 0$, customers prefer to buy old products and purchase new products when old products are out of stock; otherwise, they buy nothing

(3) When $0 \geq U_{mn}^t \geq U_{mo}^t$ or $0 \geq U_{mo}^t > U_{mn}^t$, customers buy nothing

Table 1: Parameter setting.

Parameter	Value
Average number of arrivals	$\lambda = 12$
Capacity on the shelf	$I = 15$
Fixed ordering cost	$c_f = 3$
Ordering cost per unit	$c_v = 3$
Inventory holding cost per unit	$c_I = 1$
Discount factor	$\gamma = 0.9$
New product quality	$\tau_1 = 20$
Old product quality	$\tau_2 = 12$

Then, after customer m makes a choice, there are three cases:

(1) A new product was purchased ($x_{mn}^t = 1$, $x_{mo}^t = 0$)

(2) An old product was purchased ($x_{mn}^t = 0$, $x_{mo}^t = 1$)

(3) Nothing was purchased ($x_{mn}^t = 0$, $x_{mo}^t = 0$)

After all of the customers have made their decisions, the demand for each product in the current period is $D^t(d_n^t, d_o^t)$ $(d_n^t = \sum_{i=1}^{N^t} x_{\text{in}}^t, d_o^t = \sum_{i=1}^{N^t} x_{io}^t)$.

3.4. Cost Model. At the beginning of period t, the retailer decides how many new products to order, q^t. If the quantity is 0, then $g^t = 0$; otherwise, $g^t = 1$. The product has a two-period shelf-life, and its salvage value is zero at the end of second period. The procurement lead time is assumed to be one, such that units ordered in period t will be available for sale as a "new" product in period $t + 1$. The retailer will also have to decide the number of old products to be carried into the next period, o_s^t, and these products will generate inventory holding costs $c_h^t = o_s^t c_I$. The total cost of the period is

$$C^t = g^t c_f + q^t c_v + c_h^t. \tag{3}$$

The current net profit for the period is

$$R^t = d_n^t p_n^t + d_o^t p_o^t - C^t. \tag{4}$$

3.5. Decision Model. Based on the previous two sections, we used the Markov decision process to model the sales process over an infinite horizon. At the beginning of period t, the retailer's states are $S^t(o_r^t, n^t)$ and actions are $a^t(q^t; o_s^t; p_n^t, p_o^t)$ $(o_s^t \leq \min(I - n^t, o_r^t), q^t \leq I)$. After all consumers make their purchase decisions in that period, the remaining old products are discarded directly because they have reached their lifetime threshold and the remaining new products become old products in the next period. The number of old products remaining at the beginning of period $t + 1$ is therefore $o_r^{t+1} = n^t - d_n^t$, and the number of new products in period $t + 1$ is the same as the previous order

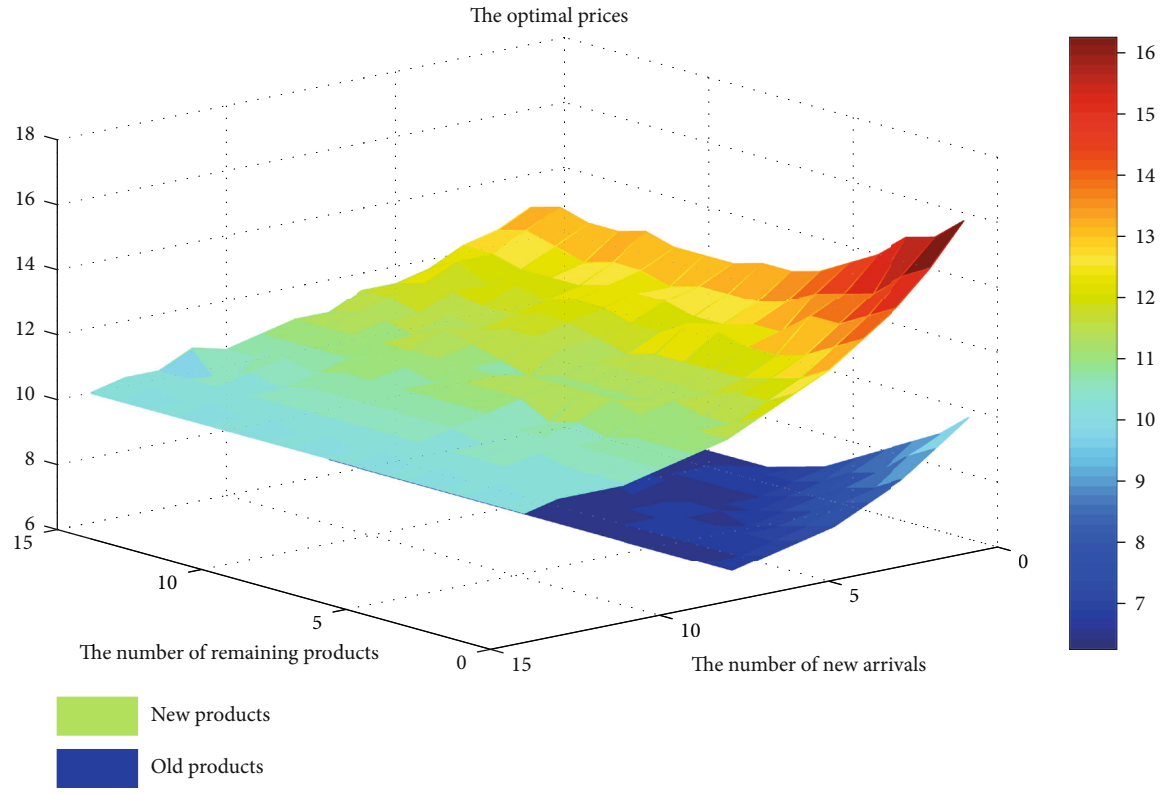


Figure 1: Basic model: optimal pricing.

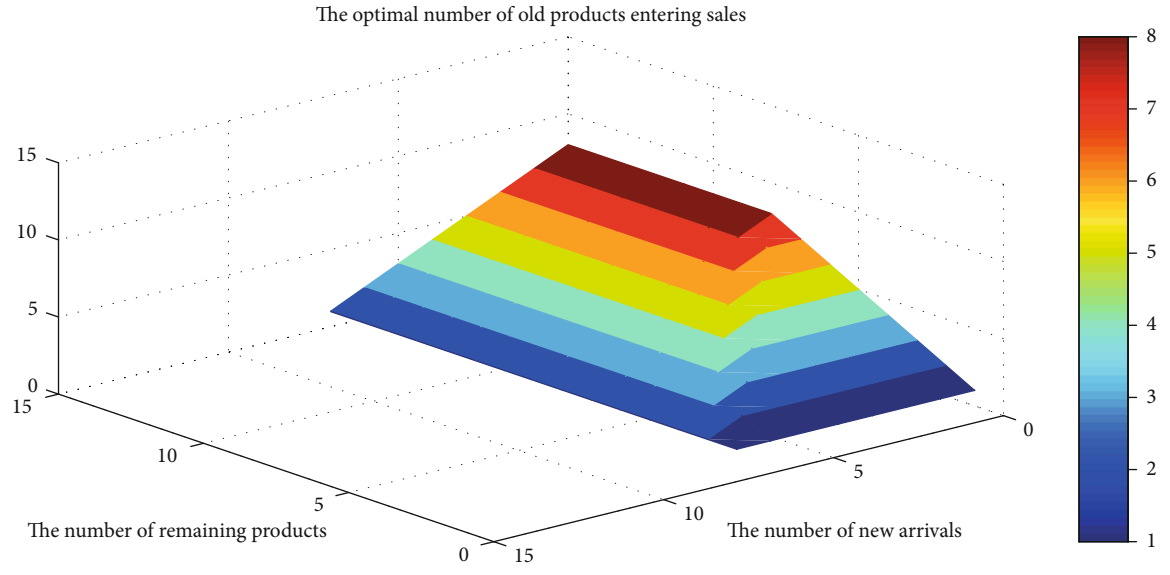


Figure 2: Basic model: the optimal number of old products carried into the next period.

quantity q^t. Therefore, the retailer's state in period $t+1$ will be $S^{t+1}(o_r^{t+1} = n^t - d_n^t, n^{t+1} = q^t)$.

Let π be a strategy that involves the actions taken in each state and Π the set of all strategies, i.e., $\pi \in \Pi$. Let V be the total discounted expected return when a certain strategy π_j is taken from a certain state S^t. Therefrom, we developed the following result:

$$V(S^t, \pi_j) = E[R^t] + \gamma E[R^{t+1}] + \gamma^2 E[R^{t+2}] + \cdots = \sum_{i=t}^{\infty} \gamma^{i-t} E[R^i]. \quad (5)$$

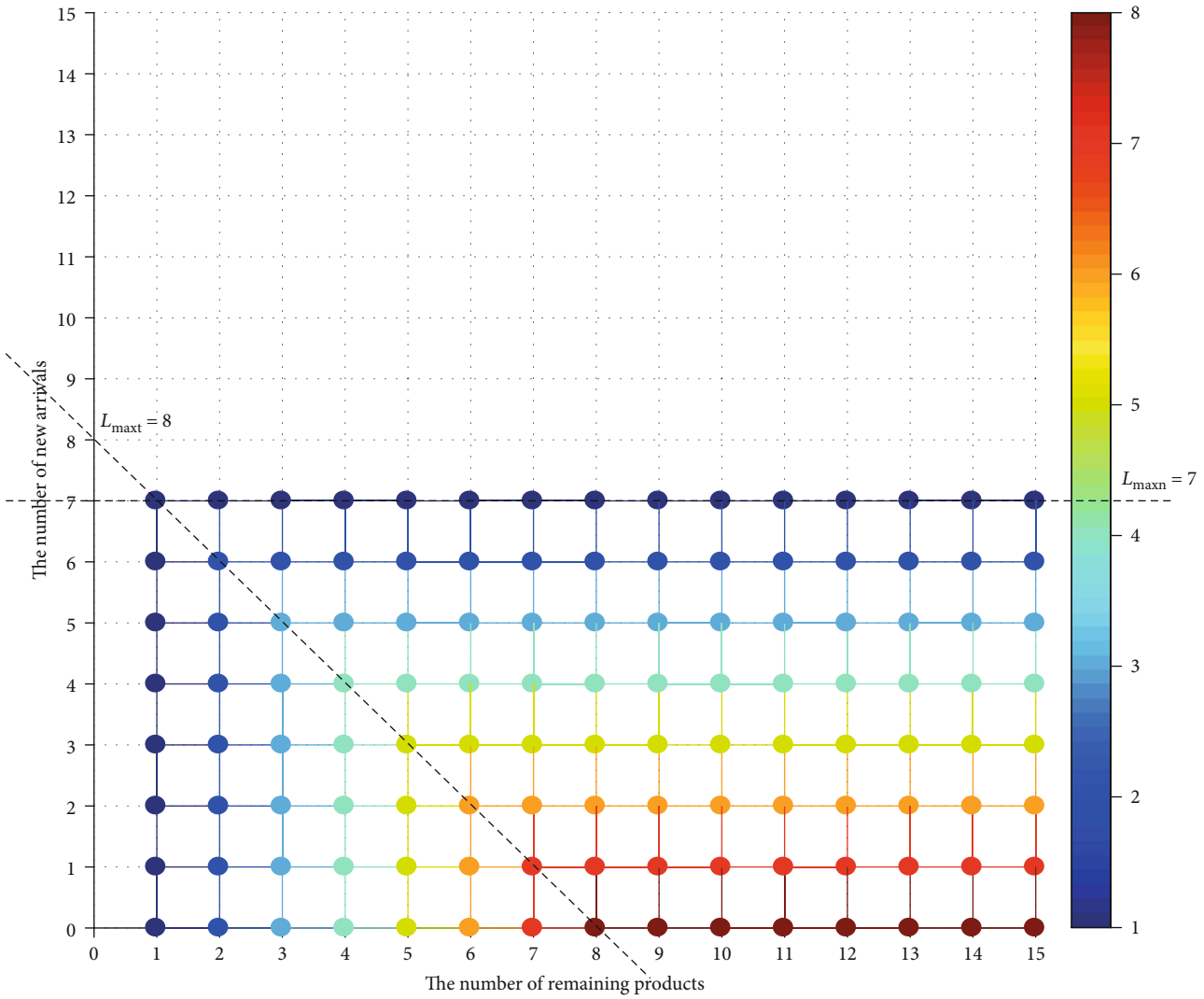


Figure 3: Basic model: the line represents the number of old products carried into the next period.

Equation (5) also has an equivalent expression, namely, the Bellman equation:

$$V\left(S^t, \pi_j\right) = E\left[R^t\right] + \gamma \bullet \sum_{S^{t+1} \in \Phi} \left[P_{S^t \to S^{t+1}, \pi_j} \bullet V\left(S^{t+1}, \pi_j\right)\right]. \quad (6)$$

The optimal strategy $\pi^*_{S^t}$ maximizes the value of state S^t, which is

$$V\left(S^t, \pi^*_{S^t}\right) = \max_{\pi \in \Pi} V\left(S^t, \pi\right). \quad (7)$$

If the policy $\pi *$ holds for all states $S \in \Phi$, then $\pi *$ is the optimal strategy for this model.

3.6. Solution Approach: Q-Learning Algorithm. In this paper, the new product orders, the number of old products carried into the next period, and pricing decisions are considered an infinite discount Markov decision model consisting of four main components: state set, action set, current expected return, and state transition probability. The states and actions of this model are $S^t(o^t_r, n^t)$ and $a^t(q^t; o^t_s; p^t_n, p^t_o)$, respectively. The states and actions for q^t and o^t_s are discrete and finite due to shelf-life constraints; furthermore, actions related to pricing can also be regarded as discrete with a price step. According to Puterman, when the state set and the action set are finite and discrete and the discount factor satisfies $0 < \gamma < 1$, the infinite discount model has an optimal

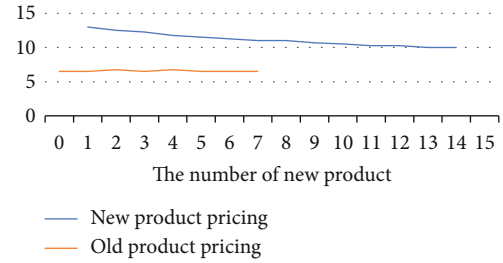


Figure 4: Optimal pricing for 2 products with 10 remaining products.

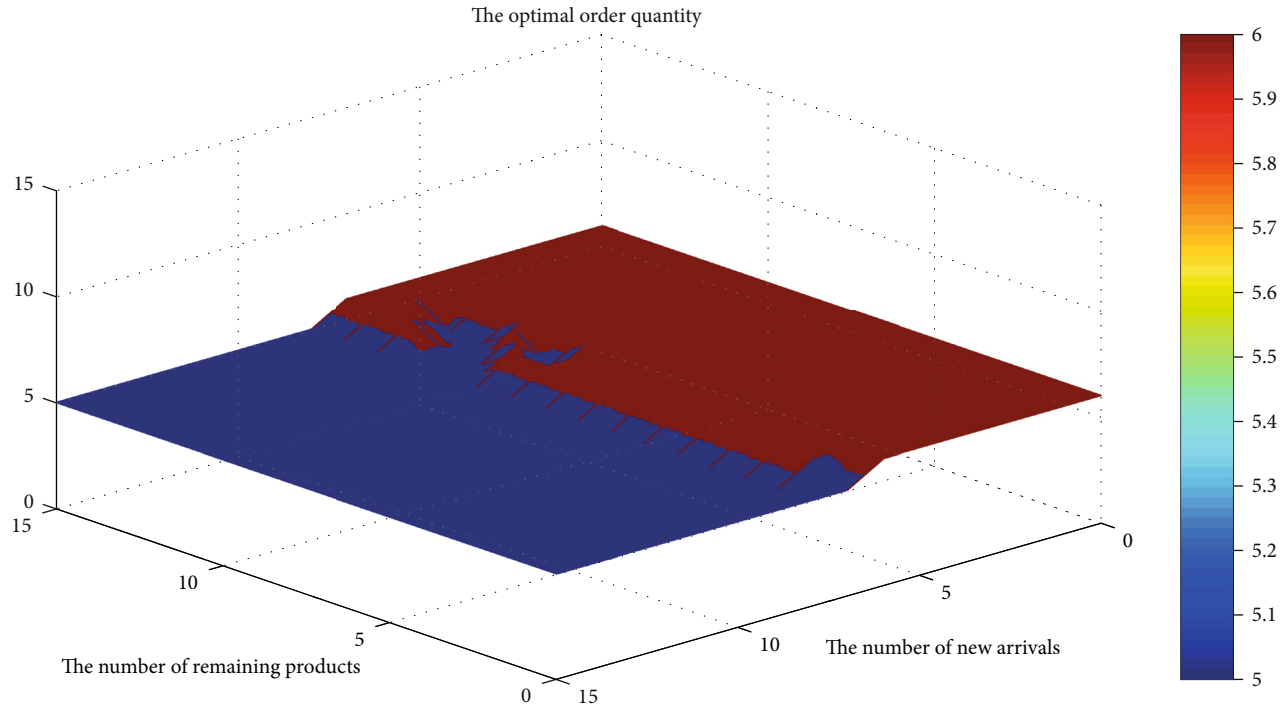


Figure 5: Basic model: optimal product order quantity.

stationary strategy [20]; in other words, the optimal strategy is only related to the states.

DP provides a number of methods for determining the optimal strategy by calculating $V*$ and one $\pi*$, assuming that R^t and $P_{S^t \to S^{t+1}, \pi_j}$ are known. One example is the value-iterative solution that Sainathan uses [18]. Although this method is intuitive, when the state action set is particularly large (for example, there are a total of 256 states in the basic model of this paper, and the largest number of possible actions exceeds 400,000), this method is very complex and unstable. Watkins and Dayan first proposed a Q-learning algorithm, which does not require knowledge of the current return and state transition probability and can determine the optimal strategy more quickly when the states and action spaces are large [21]. More recently, scholars have further optimized the Q-learning algorithm to provide precise solutions to more complex problems. Liang et al. used the deep Q-learning (DQN) algorithm to generate the watermarked positions adaptively and make some inspiration on the copyright protection issue of intellectual property (IP) circuit resources of the electronic devices in IoT environment [22]. Zhou et al. proposed an improved anisotropic Q-learning routing algorithm which can provide stable and dynamic solutions for AGV routing [23]. In recent years, Q-learning algorithm also has a good application in the field of operations management. Dittrich and Fohlmeister based on deep Q learning developed a self-optimizing inventory control, in which input is modeled as a state vector that describes the current stocks and orders within the process chain, and the output represents a control vector that controls orders for each individual station [24]. In this study, we will develop

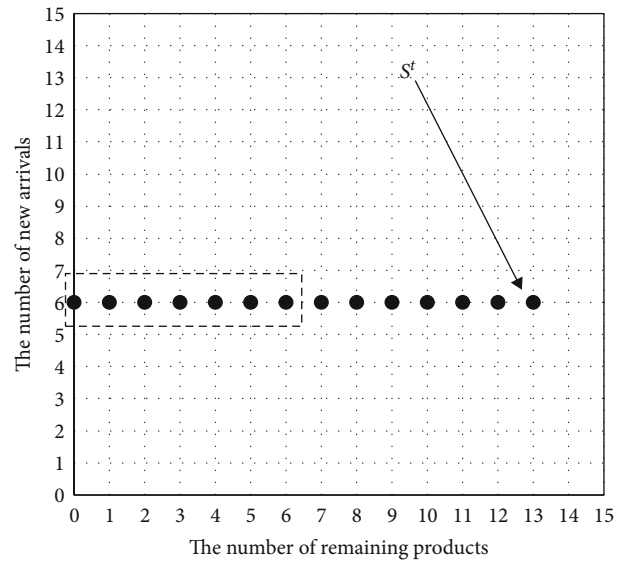


Figure 6: State transition.

the Q-learning algorithm to find a joint optimal strategy for perishable products.

Reinforcement learning methods evolved from animal learning and parameter perturbation adaptive control theory. The fundamental premise is that if an agent's actions lead to a positive environmental reward (enhanced signal), then the agent will be more likely to take this action again in the future. Q learning is an important algorithm in reinforcement learning. It combines dynamic programming with learning psychology to make the sequential optimization

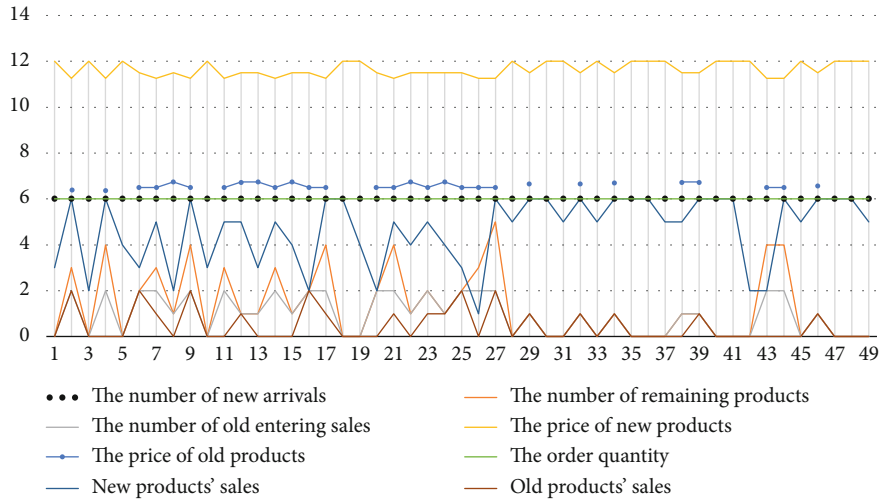


Figure 7: Parameters of the optimal strategy.

decision with delayed returns. The purpose of the reinforcement learning system is to obtain a strategy π, such that the total discount returns $\sum_{i=0}^{\infty}\gamma^i r_{t+i}$ will be maximized ($0 < \gamma < 1$ is a discount factor, γ indicates the degree of hyperopia of the system; if the value is small, then the system pays more attention to the recent actions; otherwise, it prioritizes future actions). When the state transition probability and current expected return function are unknown, the learning system can gain the optimal strategy by only using information contained in the training data for the immediate reward R^t. The learning system usually does not estimate the environment model; rather, it directly optimizes an iteratively calculated Q function. Given a strategy π, Watkins and Dayan defined the Q function as the mathematical expectation of the total discount reward when the state is s_t, the action is a_t, and the subsequent strategy is π [21].

$$Q^{\pi}(s_t, a_t) = R(s_t, a_t) + \gamma \sum_{s_t' \in S} P_{s_t s_t'}[\pi(s_t)] V^{\pi}\left(s_t'\right). \tag{8}$$

The Q value is the expected discounted reward for executing action a_t at state s_t and following policy π thereafter. The object in Q learning is to estimate the Q values for an optimal policy. It is straightforward to show that $V^*(s) = \max_{a \in A} Q^*(s, a)$ and that if $a*$ is an action at which the maximum is attained, then an optimal policy can be defined as $\pi^*(s) = a^*$. In Q learning, the agent's experience consists of a sequence of distinct stages or episodes. In the t^{th} episode, the agent observes its current state s_t, selects and performs an action a_t, observes the subsequent state s_t', receives an immediate reward r_t, and adjusts its Q_{t-1} values using a learning factor β_t according to the following.

$$Q_t(s, a) = \begin{cases} (1 - \beta_t) Q_{t-1}(s, a) + \beta_t \left[r_t + \gamma V_{t-1}\left(s_t'\right) \right] & \text{if } s = s_t \text{ and } a = a_t, \\ Q_{t-1}(s, a), & \text{otherwise,} \end{cases} \tag{9}$$

where

$$V_{t-1}\left(s'\right) = \max_b \left\{ Q_{t-1}\left(s', b\right) \right\} \tag{10}$$

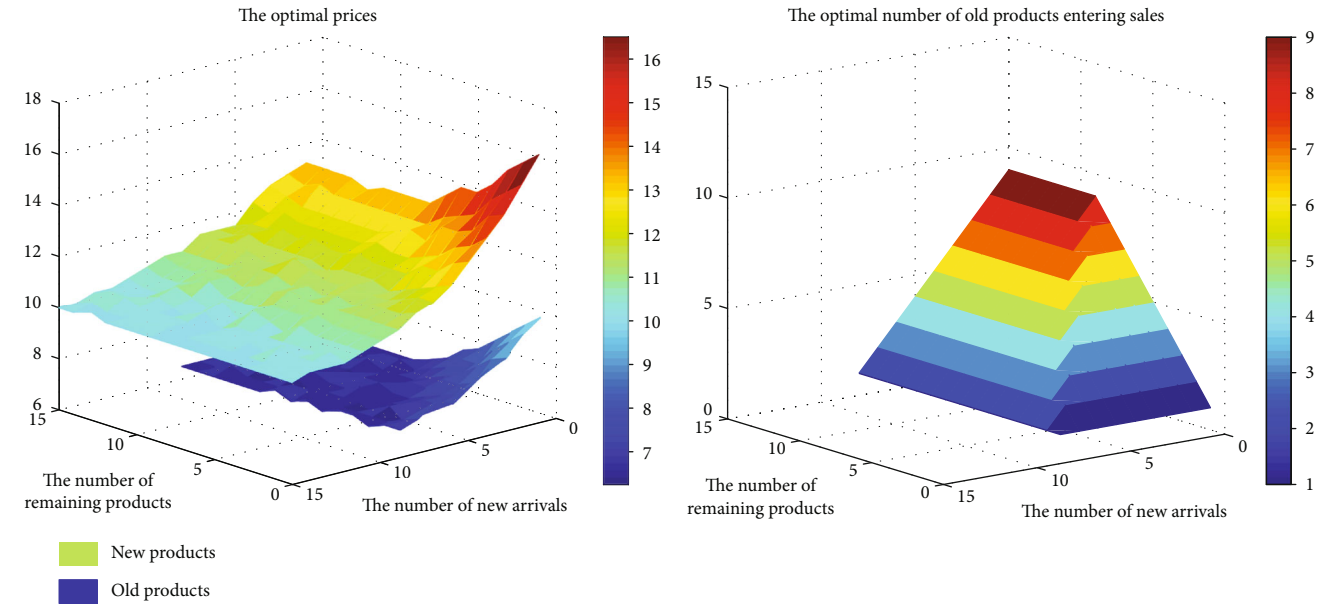
is the best the agent thinks it can do from state s'. The initial Q values, $Q_0(s, a)$, for all states and actions are assumed to be given. In Equation (9), β_t is the learning factor, which controls the speed of learning. Larger β_t indicate faster convergence. However, excessively large β_t may lead to nonconvergence. Watkins and Dayan showed that if a pair (s, a) can perform infinite iterations using Equation (9) under certain conditions, then when $t \to \infty$, $Q_t(s, a)$ has probability 1 of convergence to $Q^*(s, a)$ [21].

4. Numerical Studies and Observations

In this section, we provide an initial set of basic parameters for the simulation and then adjust parameters to analyze select features of the optimal policy. Through many numerical experiments, it is found that there are some regular patterns that can reflect the intrinsic mechanism of the model. Due to the limited space, only a set of representative parameters are selected for specific analysis and discussion. It is worth noting that the parameter settings in different situations are determined by combining the actual and the previous analysis principles to ensure the scientific results. The simulation calculation is based on the VC program and runs on the Windows platform.

4.1. Experiment Design. Table 1 shows the basic parameters used in the simulation. To ensure the accuracy of our results, we set the price step to 0.25. Finally, the average number of Q function iterations per model reached 1.5 billion and was eventually stabilized.

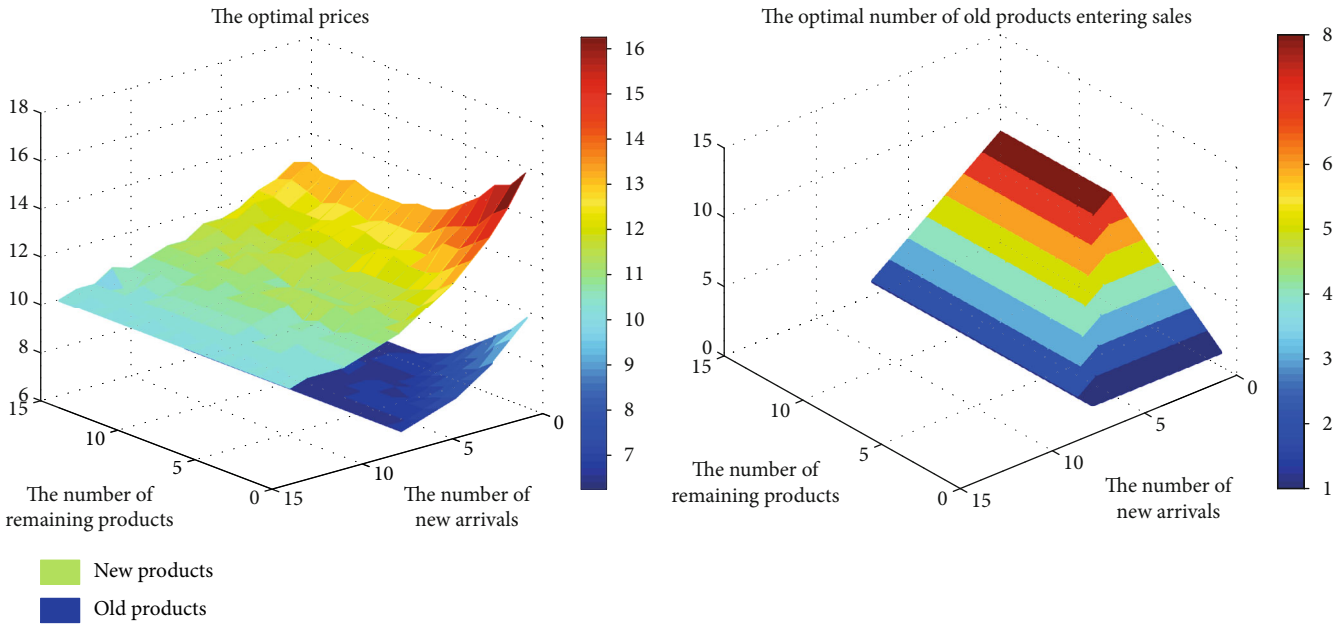
Through the computer simulation calculation, the optimal pricing strategy, the optimal number of old products carried into the next period, and the optimal ordering strategy of the basic model are given as Figures 1 and 2.



(a) The optimal pricing

(b) The optimal old product sales volume

FIGURE 8: The optimal pricing and old product sales volume with an average arrival rate of 14.



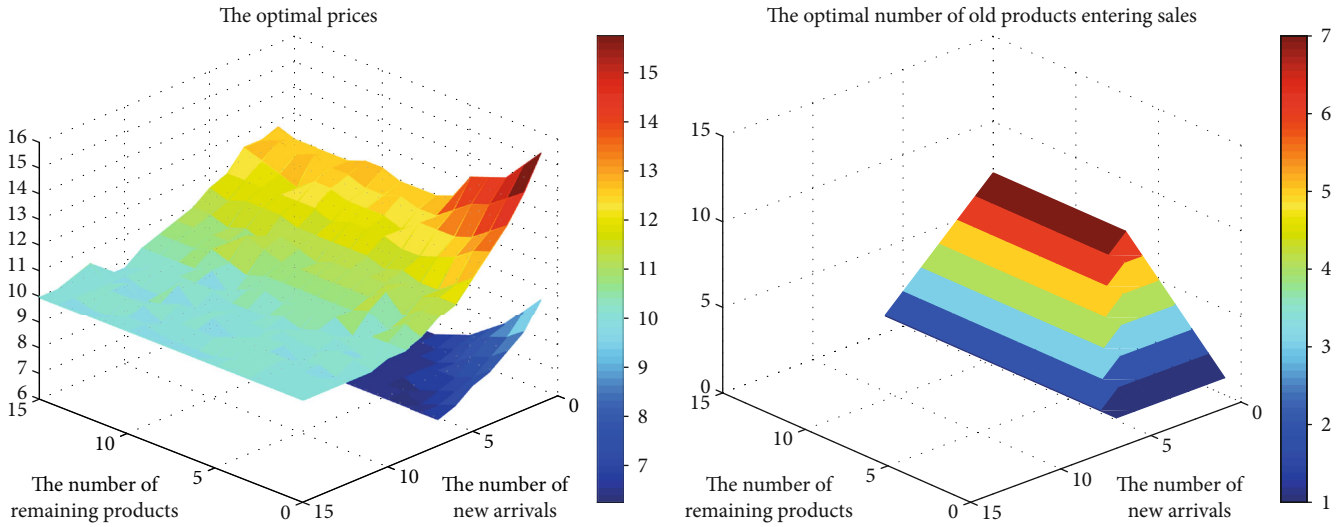
(a) The optimal pricing

(b) The optimal old product sales volume

FIGURE 9: The optimal pricing and old product sales volume with an average arrival rate of 12.

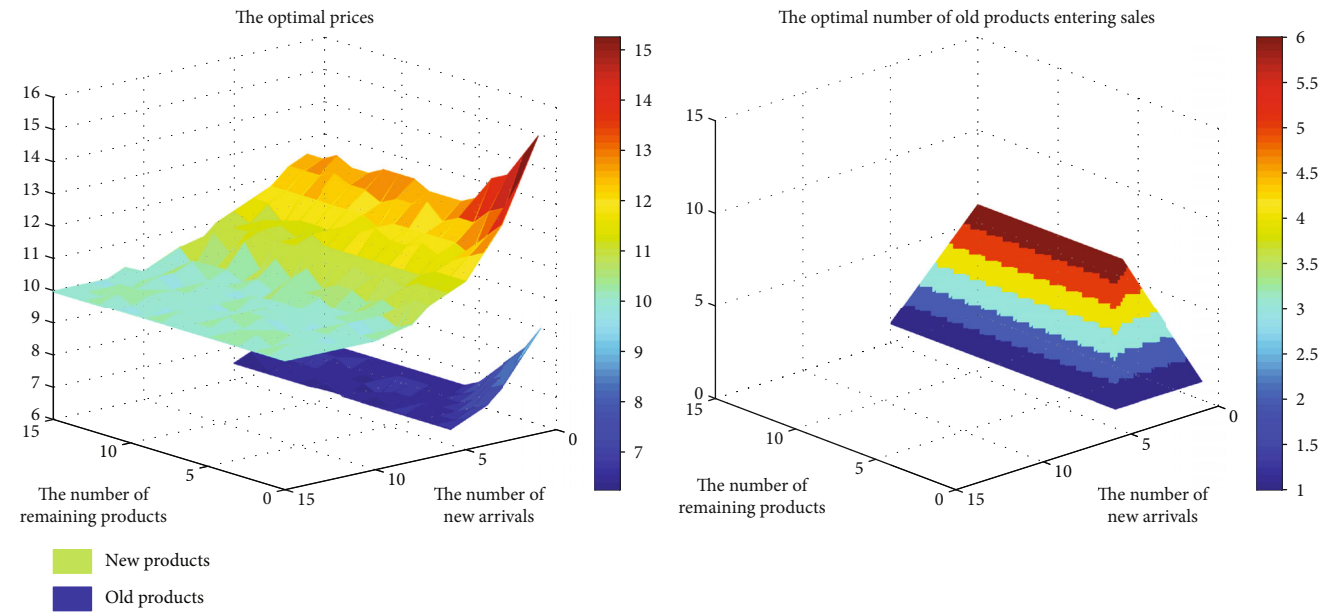
Both figures' characteristics suggest that there is an association between the optimal pricing and the number of old products carried into the next period.

As shown in Figure 3, in the region where $\{\text{new} + \text{old} \leq (L_{\max\, t} = 8)\}$ (new represents the number of new arrivals and old represents the number of remaining products), the number of old products carried into the next period is not affected by the number of new products; as such, the remaining old products all carry into the next period. In the area where $\{\text{new} + \text{old} > (L_{\max\, t} = 8), \text{new} \leq (L_{\max\, n} = 7)\}$, the number of old products carried into the next period increases as the number of new products decreases until the number of new products exceeds eight, after which no old products are sold. Therefore, the more intuitive conclusion is that there is a product bound $(L_{\max\, t})$. The retailer can yield greater profit by adding the remaining old products to the shelf when the number of new products does not exceed $L_{\max\, t}$. The findings also show that new products will be sold preferentially before old products when the potential demand is limited. Moreover, the sizes of $L_{\max\, t}$ and $L_{\max\, n}$



(a) The optimal pricing (b) The optimal old product sales volume

FIGURE 10: The optimal pricing and old product sales volume with an average arrival rate of 10.



(a) The optimal pricing (b) The optimal old product sales volume

FIGURE 11: The optimal pricing and old product sales volume with an average arrival rate of 8.

affect performance across contexts (system parameters), as we explain in greater detail in the following sections.

Based on our analysis of correlations between the optimal pricing and the optimal number of old products carried into the next period, we were able to draw some initial conclusions about the optimal pricing of both new and old products.

(1) For new products, there are four conditions affecting optimization:

First, when $\{\text{new} + \text{old} \leq (L_{\max\, t} = 8)\}$, the relationship between the new product's optimal price and the number of new and old products is negative.

Second, when $\{\text{new} + \text{old} > (L_{\max\, t} = 8)\}$, the new product's optimal price only decreases slightly as new arrivals increase, but there is no relationship to the number of old products remaining. In this region, the number of old products that get carried into the next period remains the same, even as the number of old products changes. In general, a reduction in substitute products will cause the product's price to rise, but we find that the product's price declines as the number of substitute products decreases; this indicates that increases to the number of new products have a greater impact on the price than the number of old products carried into the next period.

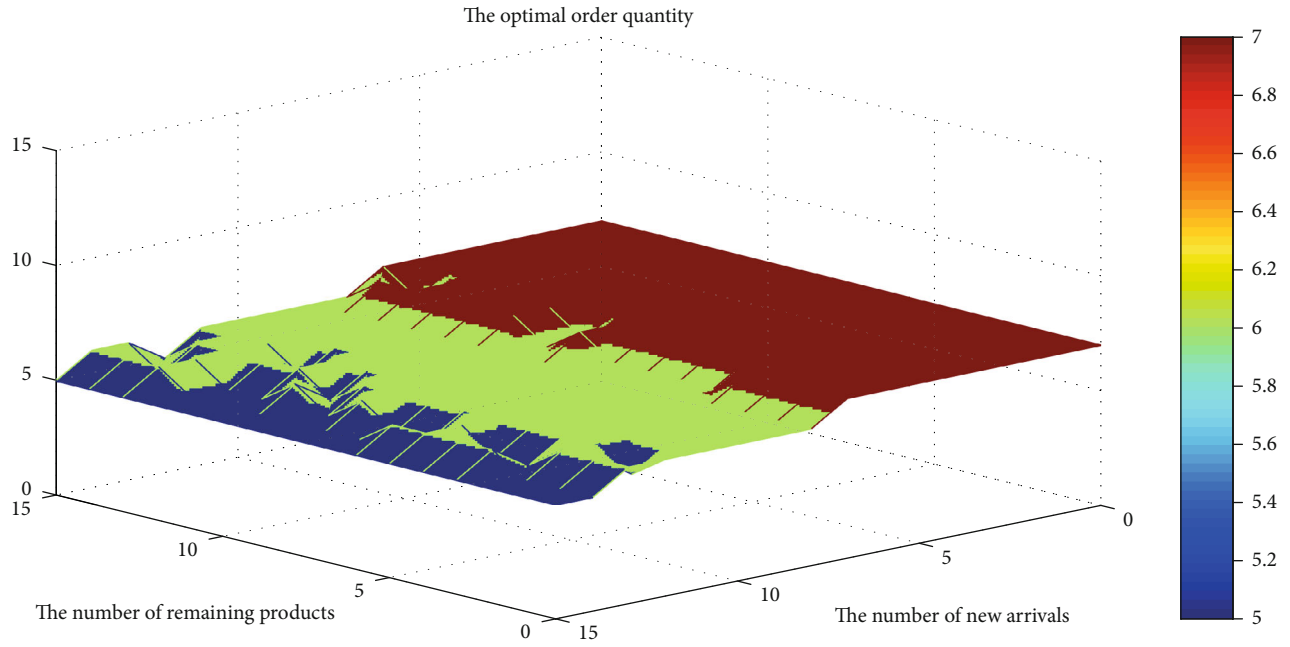


Figure 12: The optimal order quantity with an average arrival rate of 14.

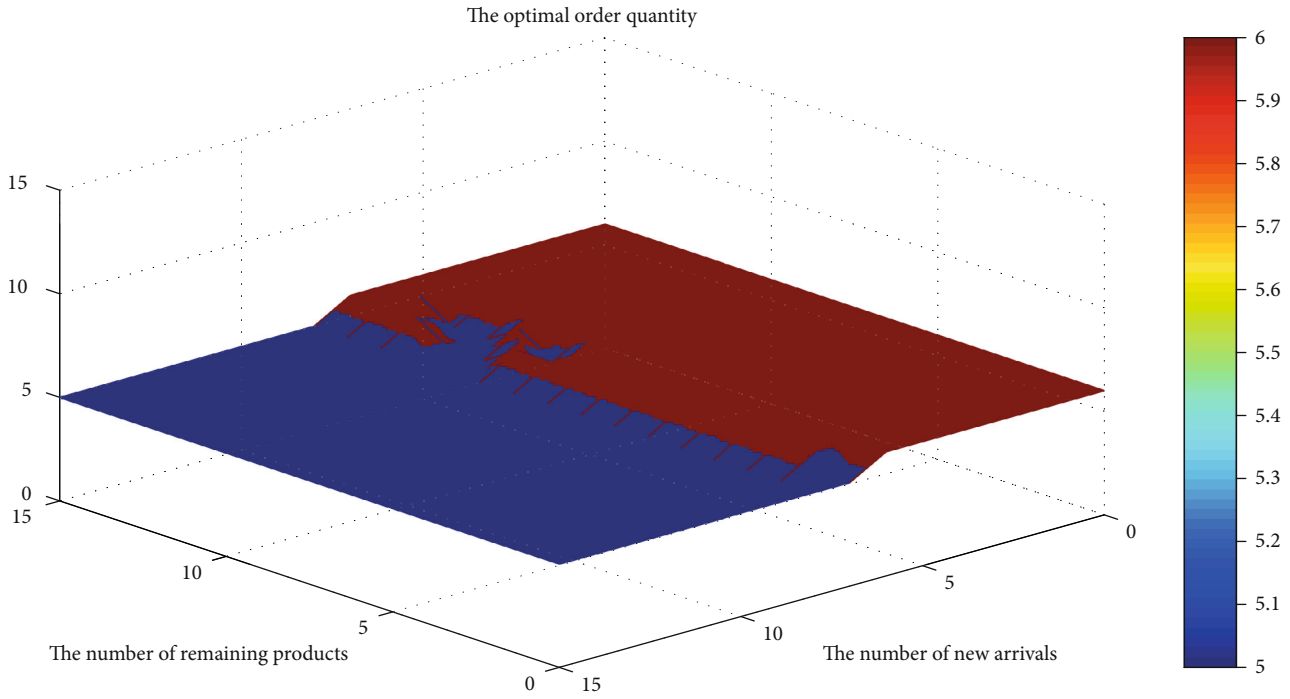


Figure 13: The optimal order quantity with an average arrival rate of 12.

Third, when $(\text{new} > 8)$, only new products are sold. Therefore, the new product price will decrease as the number of new products increases.

Finally, when $(\text{new} > 12)$, the optimal price of new products remains basically unchanged. When the sales volume approaches or exceeds the potential average number of arrivals, no more benefits can be accrued. As such, we can discern that discounts are not always beneficial.

(2) For old products, when $\{\text{new} + \text{old} \leq (L_{\max t} = 8)\}$, there is a negative relationship between the optimal price, the number of products carried over into the next period, and new product arrivals. But outside of this region, there are no obvious relationships: the old product pricing curve is generally smooth. As with new products, the pricing of old products will be affected both by the quantity of old products and

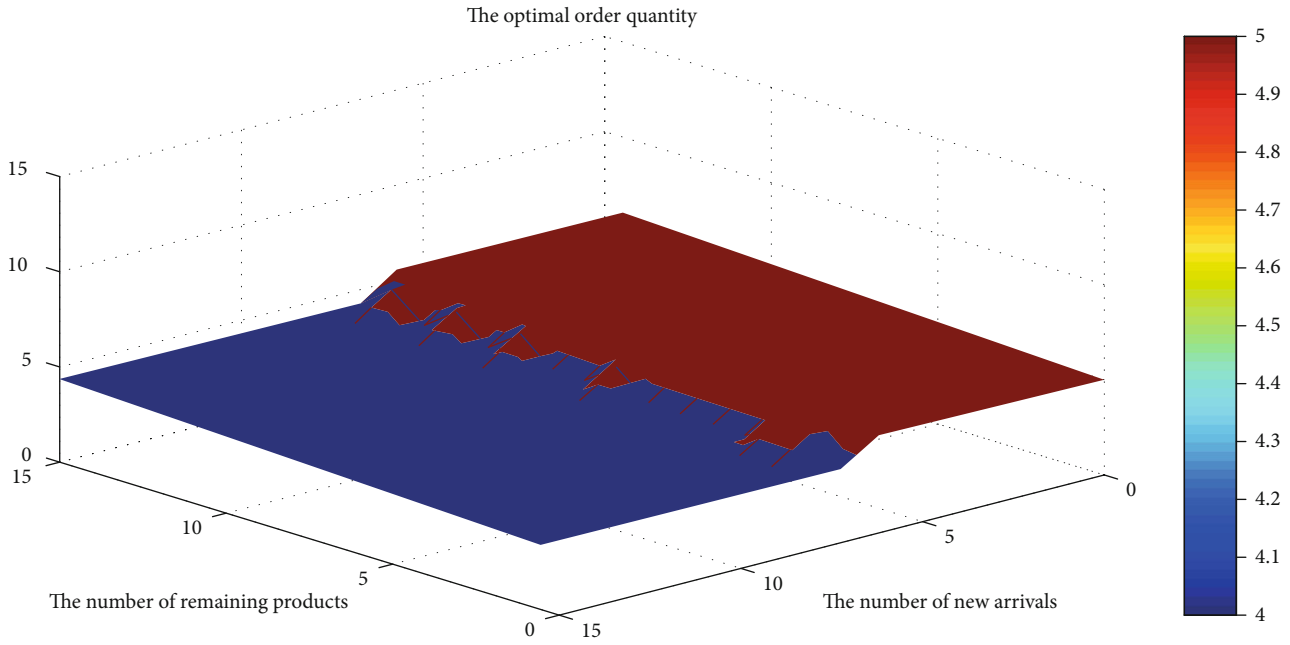


Figure 14: The optimal order quantity with an average arrival rate of 10.

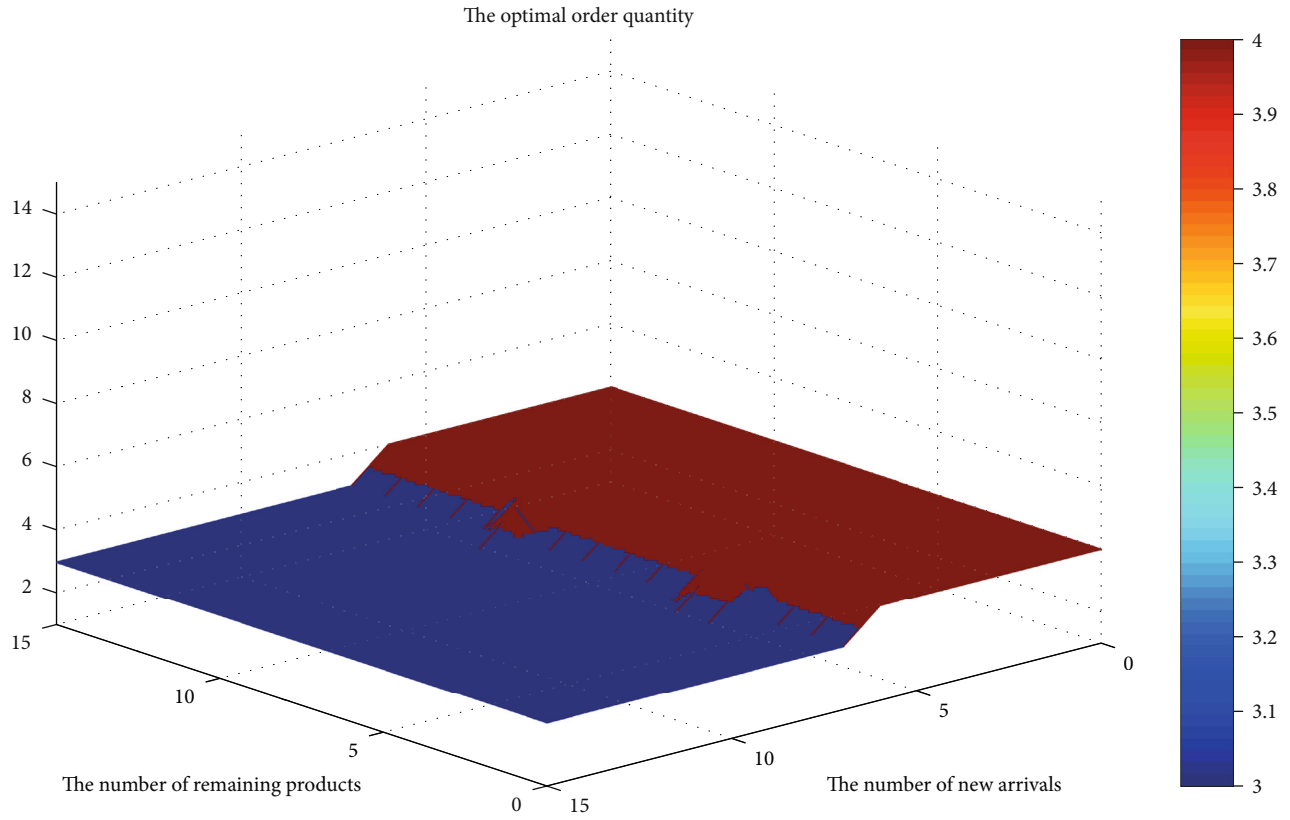
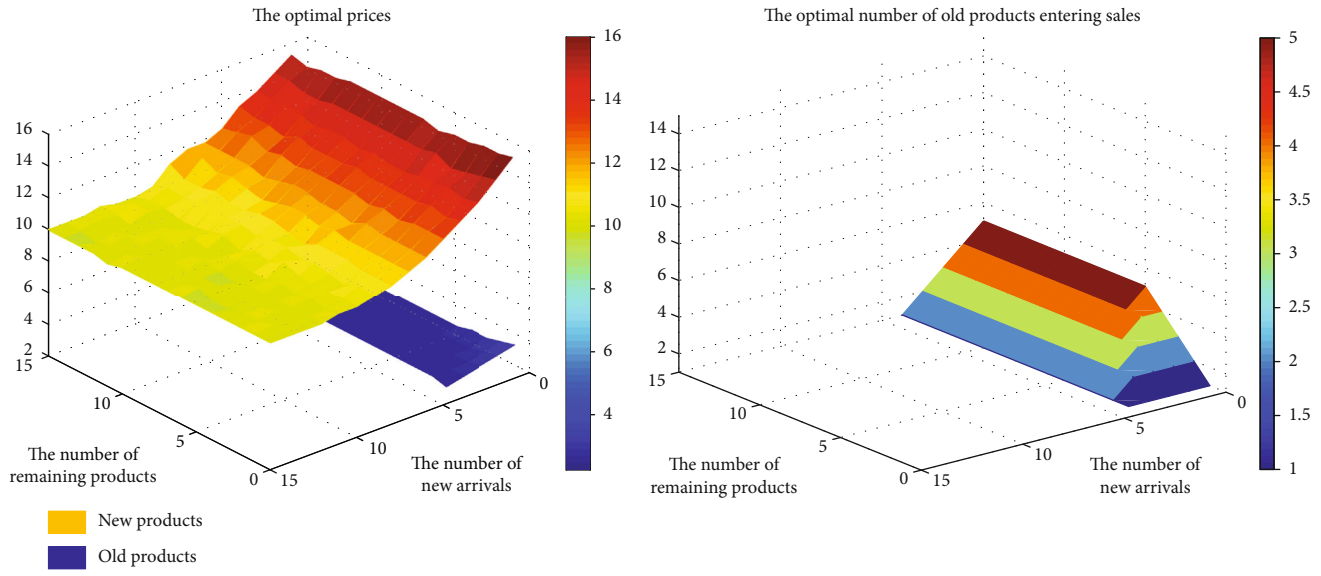


Figure 15: The optimal order quantity with an average arrival rate of 8.

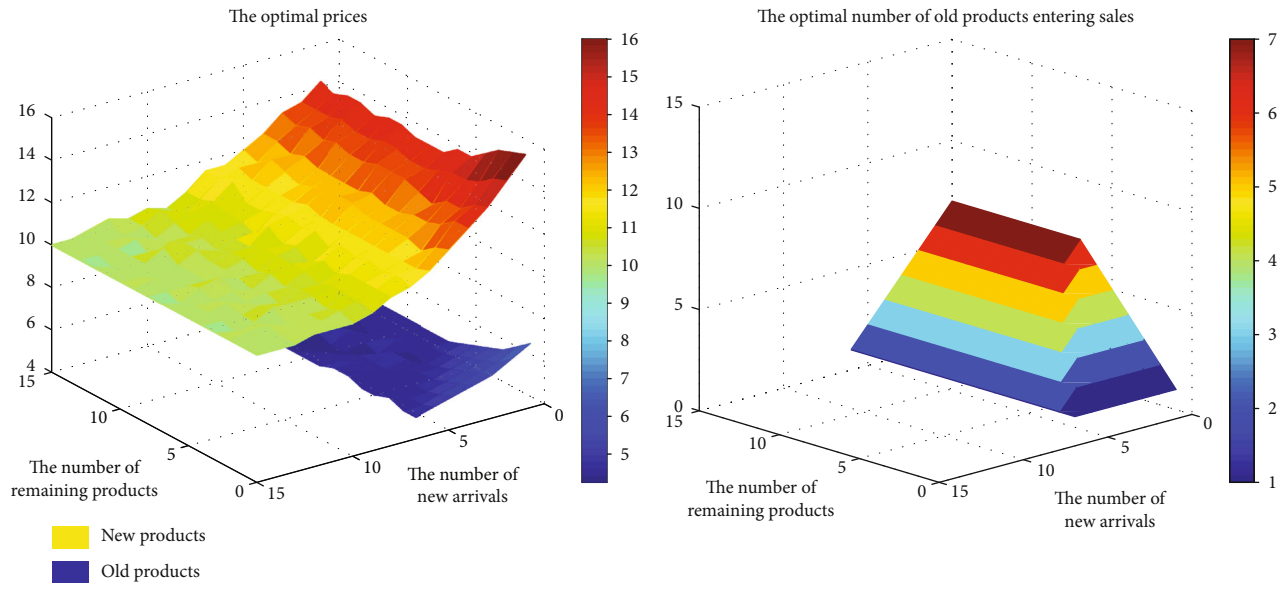
by the number of substitute products. As such, the result for this parametric setting suggests that the impacts of quantity and pricing are the same. We can therefore infer that new product quantity has a greater effect on the price of new products than on the price of old products, as shown in Figure 4



(a) The optimal pricing

(b) The optimal old product sales volume

Figure 16: The optimal pricing and old product sales volume with an 80% decay.



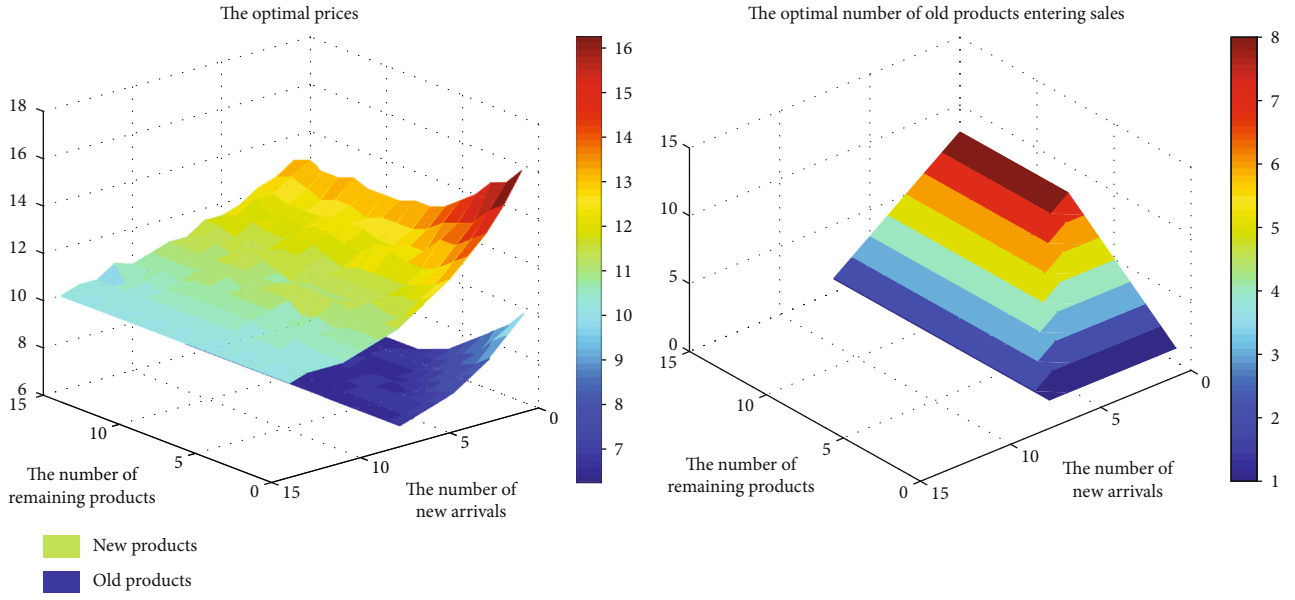
(a) The optimal pricing

(b) The optimal old product sales volume

Figure 17: The optimal pricing and old product sales volume with a 60% decay.

Figure 5 shows the optimal order quantity and reveals that new product order quantity has no relationship to the quantity of remaining products. In the $(\text{new} \leq 7)$ area, the optimal order quantity is 6, and in the $(\text{new} > 8)$ region, the optimal order quantity is 5, indicating that the expected number of old products carried into the next period will increase when there are too many new products.

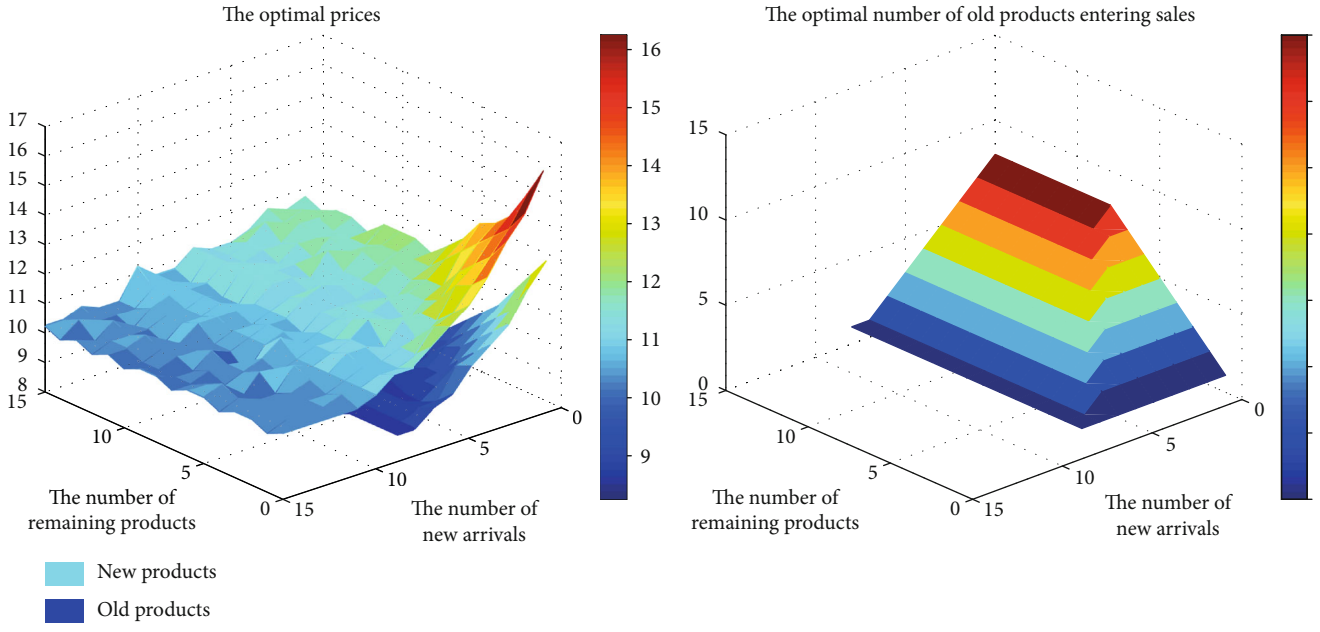
Combined with the optimal number of old products carried into the next period, the retailer faces a trade-off between decreasing order quantity to reduce ordering costs and increasing quantity to accrue a greater share of profit. Through derivation, we find that within a certain range, the probability distribution for different combinations of products can vary significantly, while the long-term expected return for different combinations may also differ. If the total number of products remains the same, the retailer can determine an appropriate ordering strategy by comparing the marginal long-term expected return between old and new product sales. However, it should be noted that long-term expected return must be considered because product combinations affect the state of the later period, which affects the long-term discounted total value of the current period.



(a) The optimal pricing

(b) The optimal old product sales volume

Figure 18: The optimal pricing and old product sales volume with a 40% decay.



(a) The optimal pricing

(b) The optimal old product sales volume

Figure 19: The optimal pricing and old product sales volume with a 20% decay.

Similarly, there is a relationship between order quantity, new product quantity, and remaining product quantity:

$$q^t = n^{t+1} \geq o_r^{t+2}. \tag{11}$$

We conducted further analysis based on this relationship. As shown in Figures 4–6, assuming that the state is S^t (new = 13, old = 9), the possible locations of the next state are shown as solid circles according to the relationship (11). After a finite number of periods, the system state is only transferred within the dashed box in the figure; further, we find that any initial state has this property.

To better understand the characteristics of optimal strategy, we simulated an optimal strategy for each period under various parameters.

As Figure 7 shows, we found the order quantity to be stable at six, but the price of new products changes within a relatively limited range, and the old product's price is

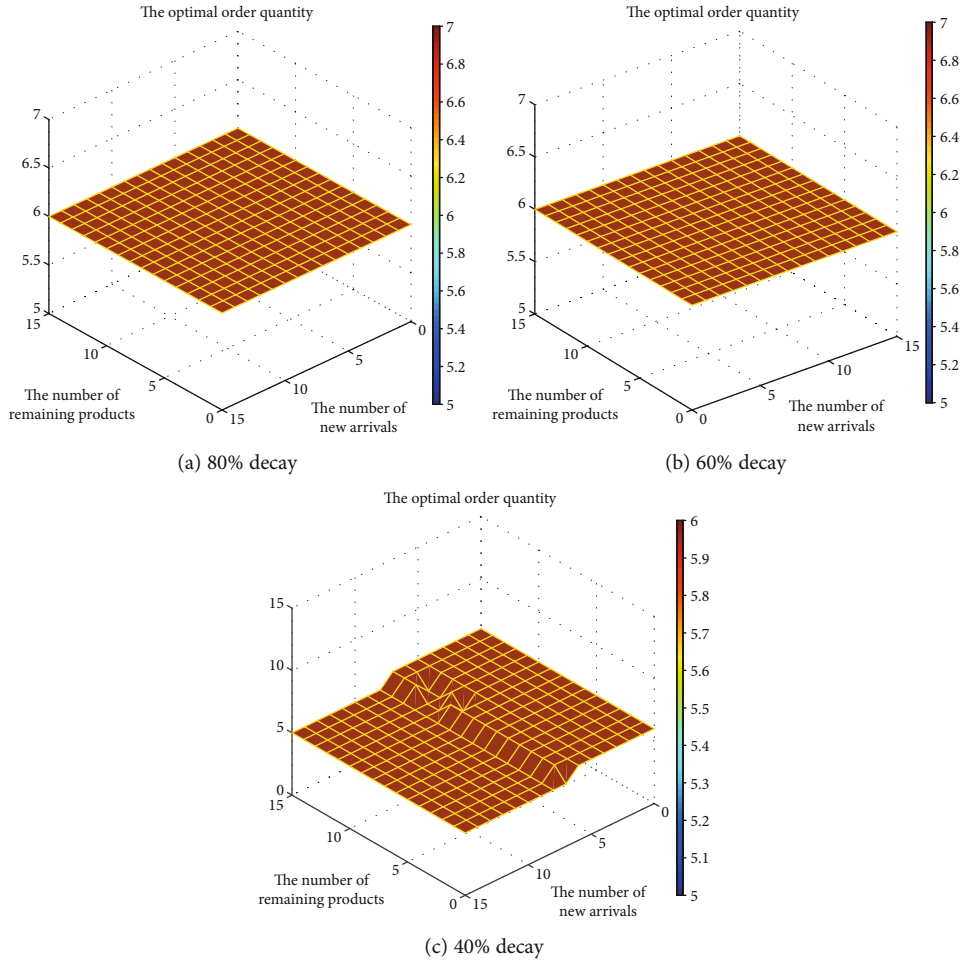


Figure 20: The optimal order quantity with 80%, 60%, and 40% decay.

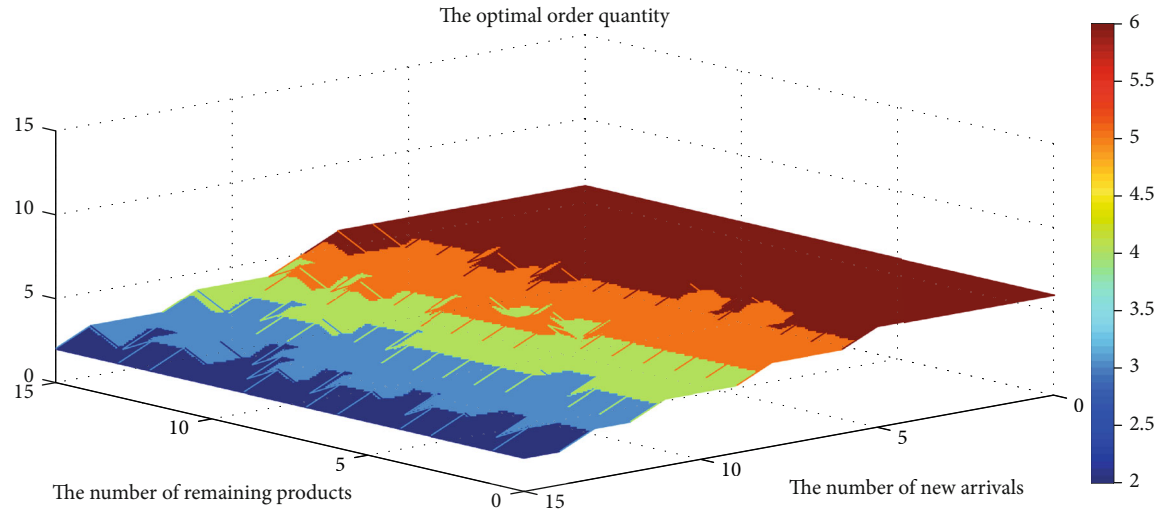


Figure 21: The optimal order quantity with a 20% decay.

also limited when sold. This suggests that (i) changes to the states and strategies are limited, which provides a basis for the retailer to enact some heuristic strategies (or fixed strategies), and (ii) even if the scope of the state transition is small, the retailer's joint strategy still presents complex features.

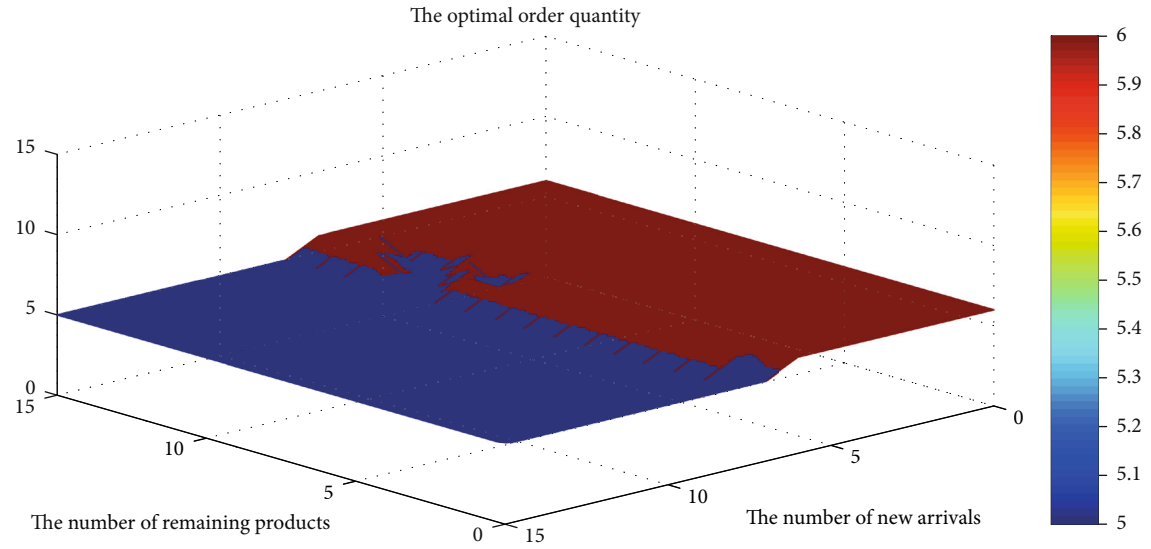


Figure 22: The optimal order quantity with an ordering cost of 3 per unit.

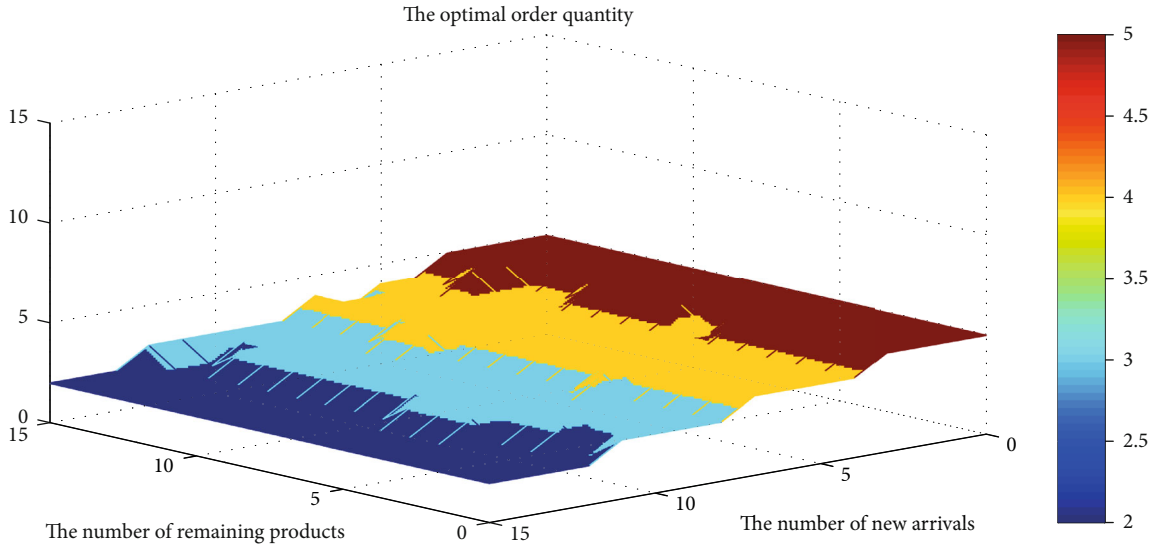


Figure 23: The optimal order quantity with an ordering cost of 5 per unit.

4.2. Expected Customer Arrivals. This section discusses the different characteristics of the optimal strategy for different expected customer arrival rates.

As shown in Figures 8–11, the price of new or old products declines as the number of customer arrivals decreases.

Likewise, the optimal order quantity decreases with the expected number of customers, as shown in Figures 12–15.

In summary, there is a positive correlation between the optimal price of both new and old products and the expected rate of customer arrivals, in addition to the optimal order quantity. For the number of old products carried into the next period, $L_{\max t}$ and $L_{\max n}$ decrease with the expected customer rate, indicating that the retailer should give the old product more shelf capacity to avoid losses when demand is adequate but prioritize new product sales if the demand is insufficient.

4.3. Product Quality Decay. Let β be the rate of perishable decay when the period is over:

$$\beta = \frac{\tau_n - \tau_o}{\tau_n}. \tag{12}$$

We can draw the following conclusion from Figures 16–19.

First, given the optimal number of old products carried into the next period in the above four figures, $L_{\max t}$ and $L_{\max n}$ increase as the decay rate decreases. A decrease in the decay rate correlates with a relative increase in the value of old products' quality such that old products have more opportunities to be sold.

Second, the optimal pricing strategy exhibits the same trend with the decay rate such that the area of the optimal price (new + old $\le L_{\max t}$) and (new $\le L_{\max n}$) expands. In this

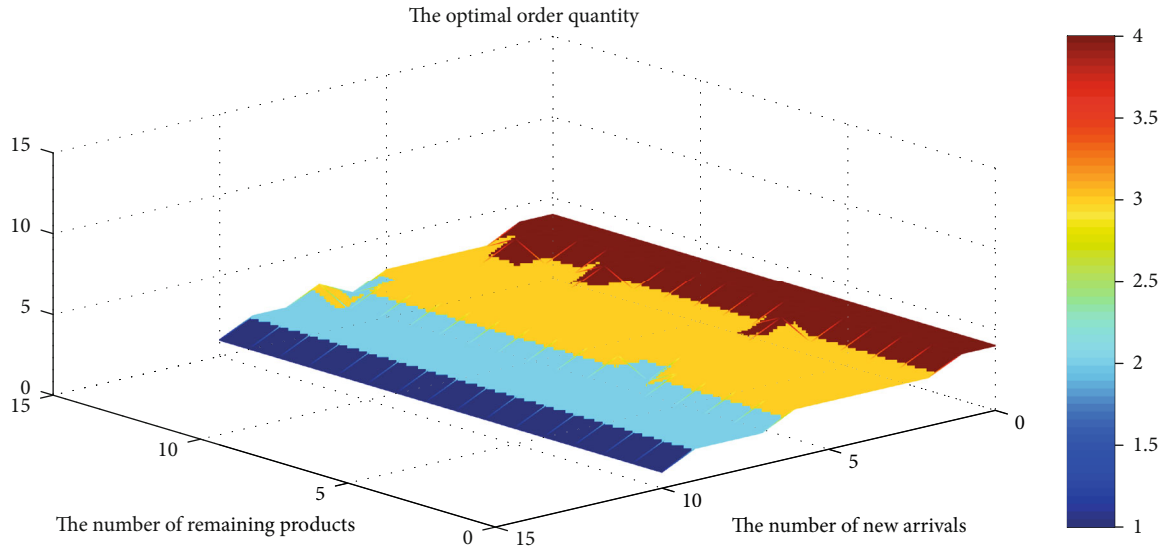


Figure 24: The optimal order quantity with an ordering cost of 7 per unit.

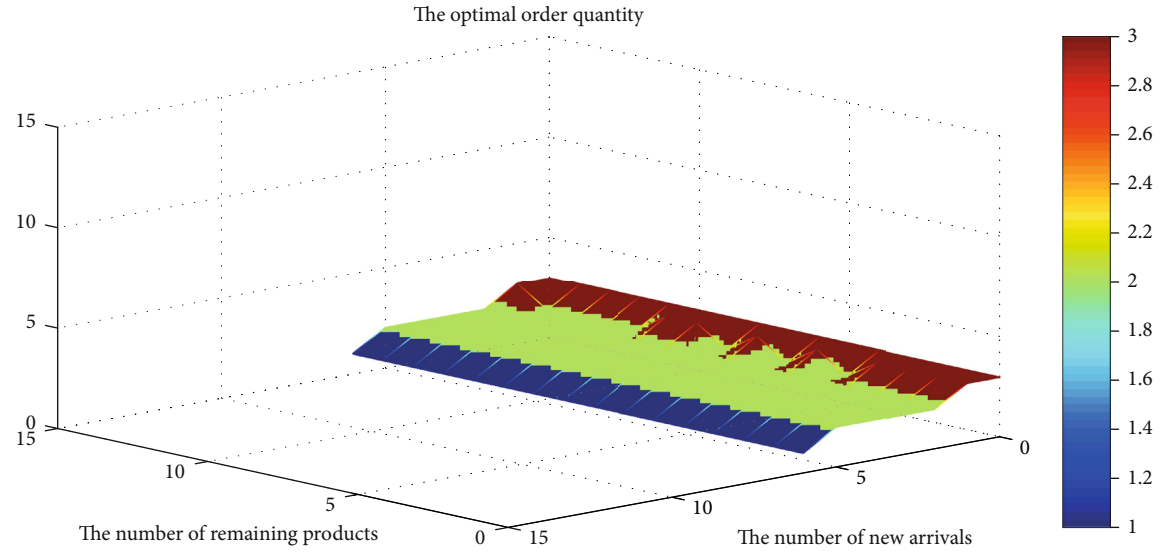


Figure 25: The optimal order quantity with an ordering cost of 9 per unit.

area, the optimal prices (for both new and old products) decrease as the number of products increases. Outside of this region, product pricing will only be affected by the number of new products because the number of old products carried into the next period decreases when the decay rate is high. In addition, the lower value of the old product has a negative effect on its own price, namely, the lower the value, the lower the price.

Figure 20 shows the optimal ordering strategy when the decay rates are 80%, 60%, and 40%, respectively. When the decay rate is either 80% or 60%, the optimal ordering strategy remains at six. However, when we decrease the rate to 40%, the characteristics are different (a detailed description accompanies Figure 5).

When we reduced the decay rate to 20%, as shown in Figure 21, the optimal ordering quantities in the $(\text{new} \geq 6)$ region decrease as the number of newly arrived products increases. This suggests that retailers will reduce current ordering when the number of new products in the current period is high and the decay rate is low.

4.4. Variable Ordering Cost per Unit. Variable ordering costs include the purchase price of the new products, the shipping costs, and other costs that are linearly related to the number of new products. This section focuses on the relationship between ordering strategies and adjustable ordering costs.

Figures 22–25 show the optimal ordering strategies when the variable costs are 3, 5, 7, and 9, respectively. First, we found the optimal ordering quantity to be unrelated to the quantity of remaining products. Second, the optimal ordering quantity decreases, exhibiting a trapezoidal decline as the variable unit cost increases. However, when the unit

order cost is too high, the possible waste due to product perishability also increases. In Figures 24 and 25, we found a case in which the ordering is 0; this indicates that when the unit order cost is too high, the retailer will try to sell the remaining old products to avoid new orders.

5. Conclusion and Future Research

Given that few works researched on pricing and inventory optimization for perishables considering multiperiod joint strategies and consumer choice behaviors, this paper conducts a simulation study centered on optimal joint strategy especially when different ages' products are sold simultaneously. In essence, this paper addresses such a problem when retailers sale a single perishable with a two-period lifetime, they should dynamically determine the joint ordering, pricing, and disposal strategy considering dynamic demand substitution, in which customers substitute between the two products when either new or old products are out of stock. In particular, the retailer sets the prices for both the new and old products and determines how many new products to order and how many remaining products to carry into the next period given demand uncertainty and diverse quality preferences among customers in each period.

We used the Markov decision process to construct the model, and the approach to seek the optimal strategy used an up-to-date version of the Q-learning algorithm. The main contribution of this paper is to utilize an accurate algorithm, Q learning, to help the retailers make their joint ordering, pricing, and disposal strategy in a time horizon allowing theoretically infinite periods. We initially apply a reinforcement learning algorithm to the inventory field, and our results also prove the efficiency of the Q learning through a large number of numerical experiments.

We further summarize the key results and insights based on our analysis as follows.

(1) We found that joint strategies of competing perishables based on dynamic ordering and pricing can yield more precise and targeted guidance for retailers. Although each strategy presents complex features, the decision-making model based on Markov decision processes has a general reference value. At the same time, determining the number of old products carried into the next period provides retailers more decision options comparing to they only decide disposal strategy, whether discard all valuable old products or enter all into the next period

(2) The optimal number of old products carried into the next period is affected by $L_{\max t}$, the total quantity of old products on shelves. All old products should be carried into the next period until the total amount of products reaches $L_{\max t}$. We found these determinations to be positively related to the potential demand and negatively related to the decay rate

(3) The pricing strategy involves several considerations: (i) optimal pricing for both new and old products is negatively related to the quantity of new and old products given that the total quantity does not exceed $L_{\max t}$; otherwise, the optimal price is only negatively related to the number of new products; (ii) the optimal pricing of both new and old products is positively correlated with the expected demand; and (iii) the optimal pricing of new products is positively related to the decay rate, but the optimal price of old products is negatively related to the decay rate

(4) Ordering decisions have complex characteristics: (i) the order size is unrelated to the quantity of remaining products but does have a positive correlation with the number of potential consumers; (ii) order quantities are not related to the number of new products when the decay rate is high or the variable ordering cost is low; however, the optimal ordering quantities exhibit a trapezoidal decline as the number of new products increases and the decay rate is low or the variable ordering cost is high

(5) In essence, our contribution to bridge the existing research gap involved both dynamic demand substitution and joint ordering, pricing, and disposal strategy for different ages' products within a multiperiod. We analyzed the optimal strategy under different parameters, by developing the Q-learning algorithm rather than dynamic programming or value iteration to solve the Markov model and gain the multiperiod optimal strategy. This provides a basis for exploring heuristic strategies and practical guidance for both academia and practice

Possible extensions of this research involve relaxing some of our assumptions, for example, by considering multiple replenishments and price changes within a period or a shelf-life of more than two periods. However, optimal solution still might be distorted due to any slight changes; as such, a more balanced measure will be expected when weighing between the model's practicality and tractability.

Data Availability

We adopted digital simulation in this paper, so the basic parameter setting is according to qualitative research of real business; there is no underlying data in this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jiangbo Zheng put forward the research topic, completed the construction of the logical framework of the paper, and finished the primary writing. Yanhong Gan wrote the program and got experimental data and finished conclusion analysis. Ying Liang gave advice and guidance on the design of the experiment and the framework of the paper and provided administrative, technical, or managerial support. Qingqing

Jiang analyzed experimental data and explored valuable experimental conclusions. Jiatai Chang had participated in paper revision and polishing.

References

[1] S. Nahmias and W. P. Pierskalla, "Optimal ordering policies for a product that perishes in two periods subject to stochastic demand," *Naval Research Logistics*, vol. 20, no. 2, pp. 207–229, 1973.

[2] B. E. Fries, "Optimal ordering policy for a perishable commodity with fixed lifetime," *Operations Research*, vol. 23, no. 1, pp. 46–61, 1975.

[3] S. Nahmias, "Optimal ordering policies for perishable inventory—II," *Operations Research*, vol. 23, no. 4, pp. 735–749, 1975.

[4] N. T. E. Morton, "Near myopic heuristics for the fixed-life perishability problem," *Management Science*, vol. 39, no. 12, pp. 1490–1498, 1993.

[5] Q. Li, P. Yu, and X. Wu, "Managing perishable inventories in retailing: replenishment, clearance sales, and segregation," *Operations Research*, vol. 64, no. 6, pp. 1270–1284, 2016.

[6] L. C. Coelho and G. Laporte, "Optimal joint replenishment, delivery and inventory management policies for perishable products," *Computers & Operations Research*, vol. 47, pp. 42–52, 2014.

[7] E. Berk and Ü. Gürler, "Analysis of the (Q, r) inventory model for perishables with positive lead times and lost sales," *Operations Research*, vol. 56, no. 5, pp. 1238–1246, 2008.

[8] X. Chao, X. Gong, C. Shi, C. Yang, H. Zhang, and S. X. Zhou, "Approximation algorithms for capacitated perishable inventory systems with positive lead times," *Management Science*, vol. 64, no. 11, pp. 5038–5061, 2018.

[9] L. Feng, "Dynamic pricing, quality investment, and replenishment model for perishable items," *International Transactions in Operational Research*, vol. 26, no. 4, pp. 1558–1575, 2019.

[10] O. Kaya and A. L. Polat, "Coordinated pricing and inventory decisions for perishable products," *OR Spectrum*, vol. 39, no. 2, pp. 1–18, 2017.

[11] M. Rabbani, N. P. Zia, and H. Rafiei, "Joint optimal dynamic pricing and replenishment policies for items with simultaneous quality and physical quantity deterioration," *Applied Mathematics and Computation*, vol. 287-288, pp. 149–160, 2016.

[12] Y. Li, A. Lim, and B. Rodrigues, "Note—pricing and inventory control for a perishable product," *Manufacturing & Service Operations Management*, vol. 11, no. 3, pp. 538–542, 2009.

[13] Y. Li, B. Cheang, and A. Lim, "Grocery perishables management," *Production and Operations Management*, vol. 21, no. 3, pp. 504–517, 2012.

[14] X. Chen, Z. Pang, and L. Pan, "Coordinating inventory control and pricing strategies for perishable products," *Operations Research*, vol. 62, no. 2, pp. 284–300, 2014.

[15] S. A. Smith and N. Agrawal, "Management of multi-item retail inventory systems with demand substitution," *Operations Research*, vol. 48, no. 1, pp. 50–64, 2000.

[16] M. E. Ferguson and O. Koenigsberg, "How should a firm manage deteriorating inventory?," *Production and Operations Management*, vol. 16, no. 3, pp. 306–321, 2007.

[17] Y. Akçay, H. P. Natarajan, and S. H. Xu, "Joint dynamic pricing of multiple perishable products under consumer choice," *Management Science*, vol. 56, no. 8, pp. 1345–1361, 2010.

[18] A. Sainathan, "Pricing and replenishment of competing perishable product variants under dynamic demand substitution," *Production & Operations Management*, vol. 22, no. 5, pp. 1157–1181, 2013.

[19] E. P. Chew, C. Lee, R. Liu, K. S. Hong, and A. Zhang, "Optimal dynamic pricing and ordering decisions for perishable products," *International Journal of Production Economics*, vol. 157, pp. 39–48, 2014.

[20] M. L. Puterman, *Markov Decision Problems*, John Wiley & Sons, 1994.

[21] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[22] W. Liang, W. Huang, J. Long, K. Zhang, K. C. Li, and D. Zhang, "Deep reinforcement learning for resource protection and real-time detection in IoT environment," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6392–6401, 2020.

[23] P. Zhou, L. Lin, and K. H. Kim, "Anisotropic Q-learning and waiting estimation based real-time routing for automated guided vehicles at container terminals," *Journal of Heuristics*, vol. 4, pp. 1–22, 2021.

[24] M.-A. Dittrich and S. Fohlmeister, "A deep q-learning-based optimization of the inventory control in a linear process chain," *Production Engineering*, vol. 15, no. 1, pp. 35–43, 2021.