



Research article

Equilibrium strategies of customers and optimal inventory levels in a make-to-stock retrial queueing system

Yuejiao Wang¹ and Chenguang Cai^{2,*}

¹ School of Mathematics and Statistics, Hunan First Normal University, 410205, Changsha, China

² School of Accounting, Hunan University of Finance and Economics, 410205, Changsha, China

* **Correspondence:** Email: ccg169@126.com.

Abstract: We examined a retrial make-to-stock system based on a double-ended queue. When the queue length was negative, the inventory system contained only products, and customers were waiting in the retrial queue when the queue length was positive. We developed a model to study the expected cost of the entire system with strategic customers in the observable case and the fully observable case. We also obtained the optimal inventory levels under these two levels of information based on numerical experiments.

Keywords: make to stock; retrial queue; double-ended; equilibrium strategies; optimal inventory level

Mathematics Subject Classification: 60K25, 91A80

1. Introduction

In this paper, we consider a make-to-stock retrial queueing system. Customers are served immediately if the system contains only products or the production server is idle upon their arrival; if the production server is busy, the customer enters a retrial queue. After the completion of service, the server searches for a customer from the retrial queue. This model reduces the cost of waiting for customers and improves the benefit to producers.

A considerable amount of research has thus been devoted to the production inventory retrial queueing system. Jose and Nair [8] examined a retrial queueing system for a production inventory under dynamic rates of production. They obtained the relevant performance measures by using the geometric matrix method based on two models, and numerically compared their expected cost functions. Jeganathan et al. [6] verified its performance by using a geometric matrix solution, and obtained the optimal inventory level through numerical simulations of a retrial inventory system. Shajin and Krishnamoorthy [16] investigated a retrial production inventory system with the (s, S)

replenishment policy. They obtained the relevant performance measures and the optimal (s, S) by using numerical simulations. In the context of an inventory retrial queueing system with multiple servers and multiple server vacations, Jose and Beena [7] obtained the performance measures of a system with two servers under the (s, S) policy, while Jeganathan et al. [5] studied its performance measures and the expected cost function. They also analyzed the performance measures and the optimal inventory level based on numerical calculations. Other studies in the area have been conducted by Reiyas and Jeganathan [15], and Jose and Reshmi [9].

Although many scholars have researched retrial queueing systems for the production inventory, most of them have examined the relevant performance measures and the optimal inventory levels under different policies, while few researchers have considered models of the production inventory with strategic customers and their influence on the optimal inventory level. Customers join the system if their individual utilities are positive, and otherwise give up trying to join it. However, every customer is influenced by other customers and seeks to attain a Nash equilibrium. Öz and Karaesmen [14] examined the optimal inventory level with strategic customers in unobservable cases. Li, Guo, Li and Song [11] studied the equilibrium strategies of customers and determined the optimal control values in an inventory system by using a negative queueing system. Zhang and Wang [22] considered a queueing system for the production inventory with a dynamic rate of production. They obtained the individual equilibrium states and the optimum cost in cases of observable as well as unobservable inventories.

Many researchers have investigated the individual and social strategic behaviors of customers in the traditional queueing system. Economous and Kanta [3] obtained the equilibrium strategies of customers and the socially optimal strategies in unobservable as well as observable cases in a retrial queueing system. Zhang [21] extended this model to a queueing system with server vacations. Wang and Zhang [18] extended it to a queueing system with dynamic rates of retrial. Wang, Zhang, and Huang [20] subsequently investigated a retrial queueing system with the N-policy. In this model, when the server was in dormant state, it did not offer any service to the customers and would not be activated until the number of customers in the wait list (retrial orbit) reached to a given threshold length N ($N \geq 1$). Wang, Liu and Zhao [19] subsequently investigated the model developed by Zhang and Huang [20] in case of multiple server vacations. Kerner and Shmuel-Bittner [10] researched the strategic behaviors of customers in a hybrid retrial queueing system. Other research in the area includes work by Li and Wang [12], Shi and Liu [17], Zhang and Wang [23], Gao, Wang, and Dong [4], and Do, Van Do, and Melikov [2].

In this study, we examine a make-to-stock queueing system with a constant retrial rate by using a double-ended queue. When the queue length is negative, the system contains only products and customers immediately take unit productions. When the queue length is positive, customers are waiting in the retrial queue. Customers arriving the system in this case find that the stock is empty and the server is busy, and then decide whether to leave their contact information or balk.

The remainder of this paper is organized as follows: In Section 2, we describe a make-to-stock retrial double-ended queueing system. Sections 3 and 4 contain discussions of the equilibrium strategies and the expected cost functions of the entire system in the observable and fully observable cases. In Section 5, we analyze the equilibrium strategies of the customers as well as their expected cost functions, and obtain the optimal inventory levels for these two cases based on numerical experiments. Section 6 summarizes the conclusions of this study.

2. Model description

We discuss a single-server make-to-stock queueing system with no waiting space. Customers arrive according to a Poisson process at rate λ . The time taken by a production server to manufacture a unit product follows an exponential distribution with rate μ . We assume that a customer takes a unit of the product once service to them has been completed. When a customer arrives, there are products in the inventory, and they take one away immediately. If the production server is serving a customer, we say that it is busy, and is otherwise idle. In order to reduce the waiting time, we stipulate that customers will not wait in the system if the production server is found to be busy. In this case, they leave their contact information and join a retrial queue according to a first-come first-serve strategy. After completion of service, the customer leaves the system and the production server searches for the next customer in the retrial queue. The search time follows an exponential distribution with rate α . When a new customer arrives during the search period, the production server stops searching and serves the new customer immediately. The server stops producing when the inventory level reaches N . The inter-arrival times of customers, and the times required for production and for the server to search for customers are mutually independent. Economou and Kanta [3] have investigated this model when $N = 0$.

$I(t)$ represents the state of the production server. $I(t) = 0$ means that the production server is idle while $I(t) = 1$ means that it is busy. $N(t)$ represents the number of customers or production units. $N(t) = -n < 0$ means that there are n units of inventory, $N(t) = n > 0$ means that there are n customers in the retrial queue, and $N(t) = 0$ means that there is neither a customer nor inventory in the system. $(I(t), N(t))$ constitutes a two-dimensional continuous-time Markov chain in state space $\{(i, n) : i = 0, 1; n = -N, -N + 1, \dots, 0, 1, \dots\}$.

A customer purchases a production unit at price f_1 , makes a profit R , and incurs a waiting cost C_1 per unit time. The production server obtains benefit f_1 when a customer purchases a production unit, and needs to pay the inventory cost C_2 per unit time.

3. The observable case

Arriving customers can observe the state of the production server in the observable case. In this case, the strategy of the customer is represented by their probability of joining the queue. When a customer finds the production server busy, they decide to join the retrial queue with probability q and balk with probability $1 - q$. When they find the production server idle, they are served immediately. Therefore, the probability of a customer joining the queue is one in this case.

Following Economou and Kanta [3], we obtain the stable condition for a make-to-stock system through Lemma 3.1.

Lemma 3.1. (*[3], Proposition 3.1*) *The necessary and sufficient conditions for the stability of the system are*

$$\rho_1 := \frac{\lambda q (\lambda + \alpha)}{\mu \alpha} < 1.$$

Let $\pi_{i,j}$ be the steady distribution of state (i, j) . We denote the partial generation function by $P_0(s) = \sum_{i=0}^{\infty} \pi_{0,i} s^i$ and $P_1(s) = \sum_{i=0}^{\infty} \pi_{1,i} s^i$. We now study the steady distribution in the almost observable case by using Lemma 3.1.

Lemma 3.2. *The steady distribution is given as follows:*

$$\begin{aligned}\pi_{0,-N} &= \frac{(1-\rho_1)(1-\rho)}{1-\rho_1+\rho_1\rho^N-q\rho^{N+1}+q\rho^{N+2}-\rho^{N+2}}; \\ \pi_{0,i} &= \rho^{i+N}\pi_{0,-N}, \quad -N+1 \leq i \leq 0; \\ P_0(1) &= \frac{1-q\rho}{1-\rho_1}\rho^N\pi_{0,-N}; \\ P_1(1) &= \frac{\rho}{1-\rho_1}\rho^N\pi_{0,-N},\end{aligned}$$

where $\rho = \frac{\lambda}{\mu}$.

Proof. We can obtain the balanced equations from Figure 1.

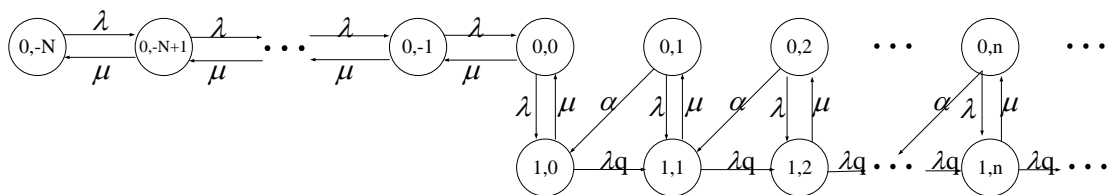


Figure 1. State transition diagram for the observable case.

$$\lambda\pi_{0,-N} = \mu\pi_{0,-N+1}; \quad (3.1)$$

$$(\lambda + \mu)\pi_{0,i} = \lambda\pi_{0,i-1} + \mu\pi_{0,i+1}, \quad -N+1 \leq i \leq -2; \quad (3.2)$$

$$(\lambda + \mu)\pi_{0,-1} = \lambda\pi_{0,-2} + \mu\pi_{0,0}; \quad (3.3)$$

$$(\lambda + \mu)\pi_{0,0} = \lambda\pi_{0,-1} + \mu\pi_{1,0}; \quad (3.4)$$

$$(\lambda + \alpha)\pi_{0,i} = \mu\pi_{1,i}, \quad i \geq 1; \quad (3.5)$$

$$(\lambda q + \mu)\pi_{1,0} = \lambda\pi_{0,0} + \alpha\pi_{0,1}; \quad (3.6)$$

$$(\lambda q + \mu)\pi_{1,i} = \lambda\pi_{0,i} + \lambda q\pi_{1,i-1} + \alpha\pi_{0,i+1}, \quad i \geq 1. \quad (3.7)$$

By (3.1), (3.2) and (3.3), we have

$$\pi_{0,i} = \rho^{i+N}\pi_{0,-N}, \quad -N+1 \leq i \leq -1,$$

$$\pi_{0,0} = \rho^N\pi_{0,-N}.$$

Multiply by s^i on both sides of (3.5) and (3.7), and summing up, we obtain

$$\begin{aligned}(\lambda + \alpha) \sum_{i=1}^{\infty} \pi_{0,i} s^i &= \mu \sum_{i=1}^{\infty} \pi_{1,i} s^i; \\ (\lambda q + \mu) \sum_{i=1}^{\infty} \pi_{1,i} s^i &= \lambda \sum_{i=1}^{\infty} \pi_{0,i} s^i + \lambda q s \sum_{i=1}^{\infty} \pi_{1,i-1} s^{i-1} + \frac{\alpha}{s} \sum_{i=1}^{\infty} \pi_{0,i+1} s^{i+1}.\end{aligned}$$

Therefore, together with (3.4) and (3.6), we have

$$(\lambda + \alpha)P_0(s) = \alpha\pi_{0,0} + \mu P_1(s),$$

$$(\lambda q + \mu)P_1(s) = \lambda P_0(s) + \lambda q s P_1(s) + \frac{\alpha}{s} P_0(s) - \frac{\alpha}{s} \pi_{0,0},$$

which yields

$$P_0(s) = \frac{\lambda q \alpha s - \alpha \mu}{\lambda q (\lambda + \alpha) s - \alpha \mu} \pi_{0,0} = \frac{1 - \rho q s}{1 - \rho_1 s} \pi_{0,0},$$

$$P_1(s) = \frac{\alpha \lambda}{\alpha \mu - \lambda q (\lambda + \alpha) s} \pi_{0,0} = \frac{\rho}{1 - \rho_1 s} \pi_{0,0}.$$

By the normalized condition,

$$\sum_{i=-N}^{-1} \pi_{0,i} + \sum_{i=0}^{\infty} (\pi_{0,i} + \pi_{1,i}) = 1,$$

we obtain

$$\pi_{0,-N} = \frac{(1 - \rho_1)(1 - \rho)}{1 - \rho_1 + \rho_1 \rho^N - q \rho^{N+1} + q \rho^{N+2} - \rho^{N+2}}.$$

□

By using the method developed in [3], we can obtain the mean waiting time of a customer who finds upon arrival that the state of the system is $(1, j)$.

$$T(1, j) = (j + 1) \frac{\lambda + \alpha + \mu}{\mu \alpha} + \frac{1}{\mu}. \quad (3.8)$$

The mean waiting time of a customer when they find the production server to be busy upon arrival can be given by

$$E(W_c | I(t) = 1) = \sum_{j=0}^{+\infty} \frac{\pi_{1,j}}{P_1(1)} T(1, j) = \frac{\lambda + \alpha + \mu}{\mu \alpha} \left(\frac{\frac{\rho \rho_1 \pi_{0,0}}{(1 - \rho_1)^2}}{\frac{\rho \pi_{0,0}}{1 - \rho_1}} + 1 \right) + \frac{1}{\mu}$$

$$= \frac{\lambda + \alpha + \mu}{\mu \alpha (1 - \rho_1)} + \frac{1}{\mu}.$$

Therefore, the utility of a customer in the almost observable case is

$$U_c(q) = R - f_1 - C_1 E(W_c | I(t) = 1) = R - f_1 - C_1 \left\{ \frac{\lambda + \alpha + \mu}{\mu \alpha (1 - \rho_1)} + \frac{1}{\mu} \right\}.$$

The equilibrium strategy of a customer in the almost observable case is specified by the joining probability. When the server is found to be busy, the customer decides to join the system with probability q_e or to balk with probability $1 - q_e$. q_e is also called the equilibrium joining strategy. If the server is found to be idle, the customer joins the system, and their probability of joining is one.

When the production server is found to be busy in the observable case, the customer's strategy can be given as follows:

Theorem 3.1. When the production server is found to be busy in the observable case, the strategy of a customer is given as follows:

$$q_e = \begin{cases} 0, & \text{if } R < f_1 + C_1 \left\{ \frac{\lambda + \alpha + \mu}{\mu \alpha} + \frac{1}{\mu} \right\}; \\ q_e^*, & \text{if } f_1 + C_1 \left\{ \frac{\lambda + \alpha + \mu}{\mu \alpha} + \frac{1}{\mu} \right\} < R < f_1 + C_1 \left\{ \frac{\lambda + \alpha + \mu}{\mu \alpha} \frac{1}{1 - \rho^*} + \frac{1}{\mu} \right\}; \\ 1, & \text{if } R > f_1 + C_1 \left\{ \frac{\lambda + \alpha + \mu}{\mu \alpha} \frac{1}{1 - \rho^*} + \frac{1}{\mu} \right\}. \end{cases}$$

where

$$q_e^* = \frac{h - 1}{h\rho^*}, \rho^* = \frac{\lambda(\lambda + \alpha)}{\mu\alpha}, h = \left[\frac{R - f_1}{C_1} - \frac{1}{\mu} \right] \frac{\mu\alpha}{\lambda + \alpha + \mu}.$$

Proof. We can then easily observe that $U_c(q)$ increases with q .

(1) If $R < f_1 + C_1 \left\{ \frac{\lambda + \alpha + \mu}{\mu \alpha} + \frac{1}{\mu} \right\}$, then $U_c(q) < 0$. Therefore, in this case, the best choice for an arriving customer is to balk, thus $q_e = 0$.

(2) If $f_1 + C_1 \left\{ \frac{\lambda + \alpha + \mu}{\mu \alpha} + \frac{1}{\mu} \right\} < R < f_1 + C_1 \left\{ \frac{\lambda + \alpha + \mu}{\mu \alpha} \frac{1}{1 - \rho^*} + \frac{1}{\mu} \right\}$, then there exists a unique solution q_e^* for $U_c(q) = 0$. q_e^* is the equilibrium joining strategy of the customer.

(3) If $R > f_1 + C_1 \left\{ \frac{\lambda + \alpha + \mu}{\mu \alpha} \frac{1}{1 - \rho^*} + \frac{1}{\mu} \right\}$, then $U_c(q) > 0$. Thus, an arriving customer's choice is to join the queue, such that $q_e = 1$. \square

We now calculate the mean cost function of the entire system.

When the queue length is negative upon a customer's arrival, their waiting time is zero; thus, the mean queue length of customers EN_c is

$$\begin{aligned} EN_c &= \lambda P_0(1) \frac{1}{\mu} + \lambda q P_1(1) E(W_c | I(t) = 1) \\ &= \frac{\rho(1 - q\rho)}{1 - \rho_1} \pi_{0,0} + \lambda q \frac{\rho}{1 - \rho_1} \pi_{0,0} \left\{ \frac{\lambda + \alpha + \mu}{\mu \alpha (1 - \rho_1)} + \frac{1}{\mu} \right\} \\ &= \frac{\rho}{1 - \rho_1} \pi_{0,0} + \frac{\rho^2 q}{(1 - \rho_1)^2} \frac{\lambda + \alpha + \mu}{\alpha} \pi_{0,0}. \end{aligned}$$

The expected inventory level EN_p is

$$EN_p = \sum_{i=-1}^{-N} (-i) \pi_{0,i} = \pi_{0,-N} \frac{(\rho^{N+1} - \rho - N(\rho - 1))}{(1 - \rho)^2}.$$

The expected cost function of the entire system $S(N)$ is

$$\begin{aligned} S(N) &= C_1 EN_c + C_2 EN_p \\ &= C_1 \frac{\rho}{1 - \rho_1} \pi_{0,0} + C_1 \frac{\rho^2 q}{(1 - \rho_1)^2} \frac{\lambda + \alpha + \mu}{\alpha} \pi_{0,0} + C_2 \pi_{0,-N} \frac{(\rho^{N+1} - \rho - N(\rho - 1))}{(1 - \rho)^2} \\ &= C_1 \frac{\rho^{N+1}}{1 - \rho_1} \pi_{0,-N} + C_1 \frac{\rho^{2+N} q}{(1 - \rho_1)^2} \frac{\lambda + \alpha + \mu}{\alpha} \pi_{0,-N} + C_2 \frac{(\rho^{N+1} - \rho - N(\rho - 1))}{(1 - \rho)^2} \pi_{0,-N}. \end{aligned}$$

This function $S(N)$ is related to the equilibrium strategy of the customer q_e^* . However, from Theorem 3.1, we know that the expression of q_e^* does not contain N . We thus use q instead of q_e^* . Our objective is to minimize the above function to obtain the optimal inventory level. However, the expected cost function is complex, because of which we analyze it by using numerical calculations.

4. Fully observable case

In this section, we study the fully observable case. In this case, customers have all the information regarding the system, then, their equilibrium strategies can be represented by the threshold n_e . When the production server is found to be idle, an arriving customer is served at once.

Let n be the threshold. When the production server is found to be busy, customers enter the retrial queue if $N(t) \leq n$, and otherwise balk.

By using the method applied by Economou and Kanta [3], we obtain the waiting time for a customer $EW_c^{ob}(n)$ when they arrive in the system, and there are n customers in the retrial queue.

$$EW_c^{ob}(n) = n \frac{\lambda + \alpha + \mu}{\mu\alpha} + \frac{1}{\mu}.$$

The utility of a customer who is in the n th position of the retrial queue is

$$U_c^{ob}(n) = R - f_1 - C_1 EW_c^{ob}(n - 1).$$

We can then obtain the equilibrium strategies of customers n_e , where this satisfies

$$n_e = \left\lceil \frac{(R - f_1)\mu\alpha}{C_1(\lambda + \alpha + \mu)} + 1 \right\rceil,$$

which satisfies

$$\begin{cases} U_c^{ob}(n_e) \geq 0; \\ U_c^{ob}(n_e + 1) < 0. \end{cases}$$

From the above function, we know that n_e is increasing because μ , α , and R are increasing. On the contrary, n_e is decreasing because λ and C_1 are increasing.

Let $\pi_{i,j}^{ob}$ be the steady distribution of state (i, j) in the fully observable case. To obtain the relevant performance measures, we investigate the steady distribution in the fully observable case as shown in Figure 2. See the following Lemma 4.1.

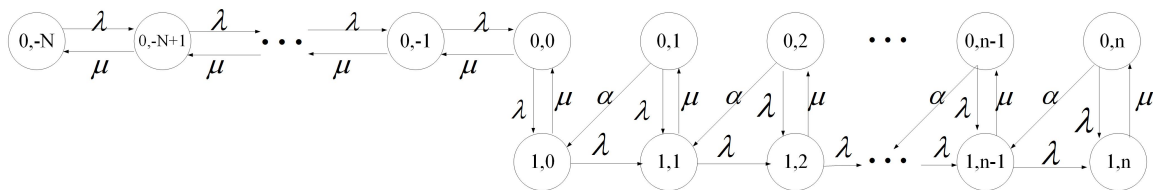


Figure 2. State transition diagram for the fully observable case.

Lemma 4.1. *The steady distribution in the fully observable case is given as follows:*

$$\begin{aligned} \pi_{0,i}^{ob} &= \rho^{i+N} \pi_{0,-N}^{ob}, -N + 1 \leq i \leq 0, \\ \pi_{0,i}^{ob} &= \beta \rho_2^i \rho \pi_{0,0}^{ob}, 1 \leq i \leq n, \\ \pi_{1,i}^{ob} &= \rho_2^i \rho \pi_{0,0}^{ob}, 0 \leq i \leq n, \\ \pi_{0,-N}^{ob} &= \frac{(1 - \rho)(1 - \rho_2)}{(1 - \rho^{N+2})(1 - \rho_2) + (\beta + 1)\rho_2\rho^{N+1}(1 - \rho_2^n)(1 - \rho)}, \end{aligned}$$

where $\beta = \frac{\mu}{\lambda + \alpha}$, $\rho_2 = \frac{\lambda(\lambda + \alpha)}{\mu\alpha}$.

Proof. The balance equations in the fully observable case are given as follows.

$$\lambda\pi_{0,-N}^{ob} = \mu\pi_{0,-N+1}^{ob}; \quad (4.1)$$

$$(\lambda + \mu)\pi_{0,i}^{ob} = \lambda\pi_{0,i-1}^{ob} + \mu\pi_{0,i+1}^{ob}, \quad -N + 1 \leq i \leq -1; \quad (4.2)$$

$$(\lambda + \mu)\pi_{0,0}^{ob} = \lambda\pi_{0,-1}^{ob} + \mu\pi_{1,0}^{ob}; \quad (4.3)$$

$$(\lambda + \alpha)\pi_{0,i}^{ob} = \mu\pi_{1,i}^{ob}, \quad 1 \leq i \leq n; \quad (4.4)$$

$$(\lambda + \mu)\pi_{1,0}^{ob} = \lambda\pi_{0,0}^{ob} + \alpha\pi_{0,1}^{ob}; \quad (4.5)$$

$$(\lambda + \mu)\pi_{1,i}^{ob} = \lambda\pi_{1,i-1}^{ob} + \lambda\pi_{0,i}^{ob} + \alpha\pi_{0,i+1}^{ob}, \quad 1 \leq i \leq n - 1; \quad (4.6)$$

$$\mu\pi_{1,n}^{ob} = \lambda\pi_{1,n-1}^{ob} + \lambda\pi_{0,n}^{ob}. \quad (4.7)$$

Substituting (4.4) into (4.6) and (4.7), we obtain

$$\begin{aligned} \pi_{1,n}^{ob} &= \rho_2\pi_{1,n-1}^{ob}; \\ (\lambda + \mu)\pi_{1,i}^{ob} &= \lambda\pi_{1,i-1}^{ob} + \frac{\lambda\mu}{\alpha + \lambda}\pi_{1,i}^{ob} + \frac{\alpha\mu}{\alpha + \lambda}\pi_{1,i+1}^{ob}, \quad 1 \leq i \leq n - 1. \end{aligned}$$

Therefore, we have

$$\pi_{1,i}^{ob} = \rho_2^i \rho \pi_{0,0}^{ob}, \quad 1 \leq i \leq n. \quad (4.8)$$

Substituting (4.8) into (4.4), we get

$$\pi_{0,i}^{ob} = \beta\pi_{1,i}^{ob} = \beta\rho_2^i \rho \pi_{0,0}^{ob}, \quad 1 \leq i \leq n.$$

By (4.1), (4.2) and (4.3), we obtain

$$\begin{aligned} \pi_{0,i}^{ob} &= \rho^{i+N} \pi_{0,-N}^{ob}, \quad -N + 1 \leq i \leq -1, \\ \pi_{0,0}^{ob} &= \rho^N \pi_{0,-N}^{ob}, \quad \pi_{1,0}^{ob} = \rho \pi_{0,0}^{ob}. \end{aligned}$$

By the normalized condition,

$$\sum_{i=-N}^{-1} \pi_{0,i}^{ob} + \sum_{i=0}^n (\pi_{0,i}^{ob} + \pi_{1,i}^{ob}) = 1,$$

we get

$$\pi_{0,-N}^{ob} = \frac{(1 - \rho)(1 - \rho_2)}{(1 - \rho^{N+2})(1 - \rho_2) + (\beta + 1)\rho_2\rho^{N+1}(1 - \rho_2^n)(1 - \rho)}.$$

□

To obtain the expected cost function of the entire system, we first give the expected queue length of customers and the expected inventory level.

According to Little's law, the expected queue length of customers is $EN_c^{ob} = \lambda_e EW_c^{ob}$, where

$$\lambda_e = \lambda(1 - \pi_{1,n}^{ob} - \sum_{i=-N}^{-1} \pi_{0,i}^{ob}) = \lambda\pi_{0,0}^{ob} \frac{(1 - \rho_2 + \beta\rho\rho_2 - \beta\rho\rho_2^n + \rho(1 - \rho_2^n))}{1 - \rho_2},$$

represents the effective rate of arrival of customers when the queue length is nonnegative, and EW_c^{ob} is the mean waiting time of a customer. When the customer arrive at the system, the state is $(0, i), i \geq 0$, and the waiting time is $1/\mu$. When the customer arrive at the system, the state is $(1, i)$, and the waiting time is $\frac{1}{\mu} + (i + 1)\frac{\lambda + \alpha + \mu}{\mu\alpha}$. Therefore, we obtain

$$EW_c^{ob} = \sum_{i=0}^n \frac{\pi_{0,i}^{ob}}{1 - \pi_{1,n}^{ob} - \sum_{i=-N}^{-1} \pi_{0,i}^{ob}} \frac{1}{\mu} + \sum_{i=0}^{n-1} \frac{\pi_{1,i}^{ob}}{1 - \pi_{1,n}^{ob} - \sum_{i=-N}^{-1} \pi_{0,i}^{ob}} \left(\frac{1}{\mu} + (i + 1) \frac{\lambda + \alpha + \mu}{\mu\alpha} \right).$$

Therefore, the mean queue length of customers is

$$\begin{aligned} EN_c^{ob} &= \lambda_e EW_c^{ob} = \lambda_e \frac{1}{\mu} + \sum_{i=0}^{n-1} \pi_{1,i}^{ob} \cdot (i + 1) \frac{\lambda + \alpha + \mu}{\mu\alpha} \\ &= \lambda_e \frac{1}{\mu} + \frac{\lambda + \alpha + \mu}{\mu\alpha} \left(\frac{\rho_2(1 - \rho_2^{n-1}) - (n - 1)\rho_2^{n-1}(1 - \rho_2)}{(1 - \rho_2)^2} \rho \pi_{0,0}^{ob} + \frac{1 - \rho_2^n}{1 - \rho_2} \rho \pi_{0,0}^{ob} \right) \\ &= \lambda_e \frac{1}{\mu} + \pi_{0,-N}^{ob} \frac{\rho^{N+1}(1 - (n + 1)\rho_2^n + n\rho_2^{n+1})}{(1 - \rho_2)^2} \frac{\lambda + \alpha + \mu}{\mu\alpha}. \end{aligned}$$

The expected inventory level EN_p^{ob} is

$$EN_p^{ob} = \sum_{i=-1}^{-N} (-i) \pi_{0,i}^{ob} = \pi_{0,-N}^{ob} \frac{(\rho^{N+1} - \rho - N(\rho - 1))}{(1 - \rho)^2}.$$

We can then obtain the expected cost function of the entire system $S_{ob}(N)$.

$$\begin{aligned} S_{ob}(N) &= C_1 EN_c^{ob} + C_2 EN_p^{ob} \\ &= C_1 \lambda_e \frac{1}{\mu} + C_1 \pi_{0,-N}^{ob} \frac{\rho^{N+1}(1 - (n + 1)\rho_2^n + n\rho_2^{n+1})}{(1 - \rho_2)^2} \frac{\lambda + \alpha + \mu}{\mu\alpha} \\ &\quad + C_2 \pi_{0,-N}^{ob} \frac{(\rho^{N+1} - \rho - N(\rho - 1))}{(1 - \rho)^2}. \end{aligned}$$

Because the expression of $S_{ob}(N)$ is complex, we determine the optimal inventory level by using numerical calculations.

5. Numerical experiments

We investigate the equilibrium strategies of customers and the expected cost functions in two cases through numerical experiments. First, we show how the parameters (such as the rates of customer arrival λ , production μ , and retrial α) influence the equilibrium strategies of the customers in two cases q_e^*, n_e^* . Second, we analyze the expected cost functions in the two cases ($S(N), S_{ob}(N)$) under the influence of the parameters of the system (λ, μ, α) . Finally, we obtain the optimal inventory level N based on values of the expected cost functions in two cases.

We assume that the profit from a customer's purchase of a production unit is $R = 15$, the cost of purchasing a production unit is $f_1 = 5$, the customers' unit waiting cost is $C_1 = 3$, and the cost of inventory for production in unit time is $C_2 = 1$.

5.1. Equilibrium strategies of customers

As shown in Table 1, the equilibrium strategies of the customers decrease as λ increases in both cases. When more customers arrive at the system, their waiting time increases such that many customers are unwilling to enter the system. The customers' strategies are increasing for μ in the two cases presented in Table 2 because a reduction in the production time reduces the customer's waiting time. We know from Table 3 that the equilibrium strategies of the customers are increasing for α in the two cases. This means that the customers are likely to receive service with a reduction in their waiting time.

In summary, we know that the equilibrium strategies of the customers decrease as λ increases, and increase with μ and λ in two cases. These results are understandable.

Table 1. Equilibrium strategies in the observable and fully observable case for $\mu = 15$, $\alpha = 4$.

λ	5.7	5.8	5.9	6	7	8	10	12	14
q_e^*	0.95	0.92	0.89	0.87	0.68	0.54	0.37	0.26	0.20
n_e^*	9	9	9	9	8	8	7	7	7

Table 2. Equilibrium strategies in the observable and fully observable case for $\lambda = 15$, $\alpha = 6$.

μ	20	21	22	25	28	34	40	44	50
q_e^*	0.34	0.36	0.38	0.43	0.48	0.59	0.70	0.78	0.88
n_e^*	10	11	11	11	12	13	14	14	15

Table 3. Equilibrium strategies in the observable and fully observable cases for $\lambda = 15$, $\mu = 40$.

α	3	4	4.5	5	6	7	8	9	9.5
q_e^*	0.38	0.5	0.55	0.61	0.70	0.79	0.87	0.95	0.98
n_e^*	7	10	11	12	14	16	17	19	20

5.2. The expected cost function of the system

We can determine from the left side of Figure 3 that the expected cost function of the entire system is a concave function of λ in both cases. As λ increases, so does the speed at which products are taken from the inventory to reduce the cost of inventory. However, when λ increases to a fixed value, the waiting time for customers increases such that the expected cost of the entire system increases. The right side of Figure 3 illustrates that the expected cost function of the entire system decreases with respect to μ in both cases. When $\mu > 1.5$, the mean cost of the system does not change by much. The waiting cost of the customers is very short when μ is large, while the cost of inventory of products remains almost unchanged. Thus, the mean cost of the entire system tends to be stable.

The left side of Figure 4 indicates that the expected cost of the system decreases and tends to stabilize as α increases in both cases. This implies that when α increases to a certain level, the expected waiting cost of customers changes by little, such that the cost of the entire system tends toward a constant value.

The right side of Figure 4 illustrates that the expected cost functions of the entire system are concave functions with respect to N in the two cases considered. Thus, there exists an optimal inventory level in our model. This level is 1 in the observable case, and 2 in the fully observable case. These results provide a sound basis for production managers to plan the levels of their inventory.

Because the expected waiting cost increases with q and n , the expected cost functions of the entire system are an increasing function of q and n in both cases, respectively—see Figure 5.

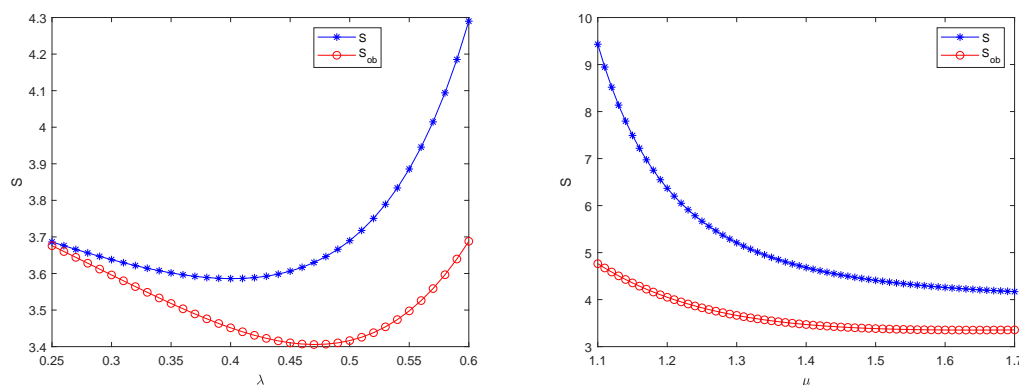


Figure 3. Left: The expected cost functions of entire system in two cases vs. λ for $\mu = 1$, $q = 0.9$, $\alpha = 4$, $N = 4$, $n = 5$; Right: The expected cost functions of entire system in two cases vs. μ for $\lambda = 0.8$, $q = 0.9$, $\alpha = 4$, $N = 4$, $n = 5$.

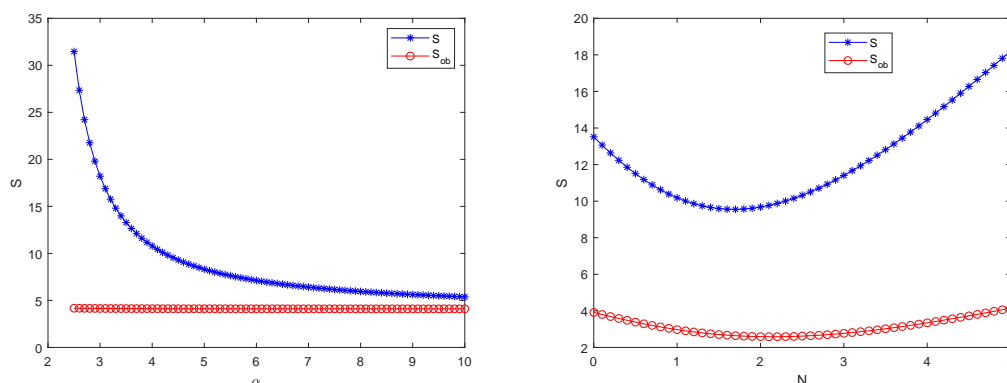


Figure 4. Left: The expected cost functions of entire system in two cases vs. α for $\lambda = 2$, $\mu = 4$, $n = 5$, $N = 5$, $q = 0.9$; Right: The expected cost functions of entire system in two cases vs. N for $\lambda = 2$, $\mu = 4$, $\alpha = 3$, $n = 5$, $q = 0.9$.

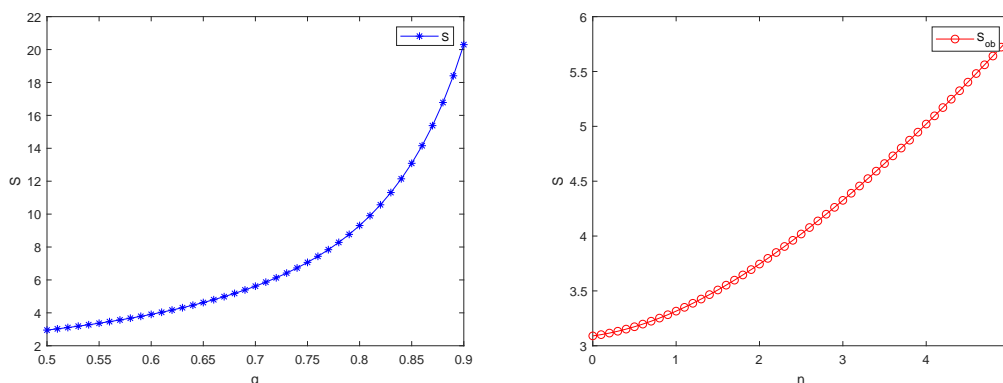


Figure 5. Left: The expected cost function of entire system in the observable case vs. q for $\lambda = 0.8, \mu = 1, \alpha = 4, N = 5$; Right: The expected cost function of entire system in the fully observable case vs. n for $\lambda = 0.8, \mu = 1, \alpha = 4, N = 5$.

6. Conclusions

In this paper, we considered a make-to-stock retrial double-ended queueing system. When customers find that the production server is busy upon their arrival, they decide whether to join the system or to balk. We discussed the equilibrium strategies of the customers and the expected cost functions of the system in two cases. Furthermore, we examined the equilibrium strategies of the customers and the expected cost functions, and calculated the optimal inventory levels under two states of information based on numerical experiments. We also obtained the optimal inventory levels for the two cases through simulations. Our results here provide a basis for managers to make decisions to optimize their inventories.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This article has been partially supported by the National natural Science Foundation of China (Grants no.11901186, 12271062), the Social Science Foundation of Hunan Province (No. 19YBQ113), Scientific Research Foundation of Hunan Education Department (No. 23A0679) and Natural Science Foundation of Hunan Province (No. 2021JJ41088).

Conflict of interest

The authors declare no conflict of interest in this paper.

References

1. G. Arivarignan, M. Keerthana, B. Sivakumar, A production inventory system with renewal and retrial demands, *Appl. Pro. Sto. Process.*, (2020), 129–142. https://doi.org/10.1007/978-981-15-5951-8_9
2. N. H. Do, T. VanDo, A. Melikov, Equilibrium customer behavior in the M/M/1 retrial queue with working vacations and a constant retrial rate, *Oper. Res. Ger.*, **20** (2020), 627–646. <https://doi.org/10.1007/s12351-017-0369-7>
3. A. Economou, S. Kanta, Equilibrium customer strategies and social-profit maximization in the single-server constant retrial queue, *Nav. Res. Log.*, **58** (2011), 107–122. <https://doi.org/10.1002/nav.20444>
4. S. Gao, H. Dong, X. Wang, Equilibrium and pricing analysis for an unreliable retrial queue with limited idle period and single vacation, *Oper. Res. Ger.*, **21** (2021), 621–643. <https://doi.org/10.1007/s12351-018-0437-7>
5. K. Jeganathan, T. Harikrishnan, K. P. Lakshmi, D. Nagarajan, A multi-server retrial queueing-inventory system with asynchronous multiple vacations, *Dec. Anal. J.*, **9** (2023), 100333. <https://doi.org/10.1016/j.dajour.2023.100333>
6. K. Jeganathan, M. A. Reiyas, S. Selvakumar, N. Anbazhagan, Analysis of retrial queueing-inventory system with stock dependent demand rate: (s, S) versus (s, Q) ordering policies, *Int. J. Appl. Comput. Math.*, **6** (2020), 1–29. <https://doi.org/10.1007/s40819-020-00856-9>
7. K. P. Jose, P. Beena, On a retrial production inventory system with vacation and multiple servers, *Int. J. Appl. Comput. Math.*, **6** (2020), 108. <https://doi.org/10.1007/s40819-020-00862-x>
8. K. P. Jose, S. S. Nair, Analysis of two production inventory systems with buffer, retrials and different production rates, *J. Indust. Eng. Int.*, **13** (2017), 369–380. <https://doi.org/10.1007/s40092-017-0191-0>
9. K. P. Jose, P. S. Reshmi, A production inventory model with deteriorating items and retrial demands, *Opsearch*, **58** (2021), 71–82. <https://doi.org/10.1007/s12597-020-00471-8>
10. Y. Kerner, O. Shmuel-Bittner, Strategic behavior and optimization in a hybrid M/M/1 queue with retrials, *Que. Syst.*, **96** (2020), 285–302. <https://doi.org/10.1007/s11134-020-09672-w>
11. Q. Li, P. Guo, C. L. Li, J. S. Song, Equilibrium joining strategies and optimal control of a make-to-stock queue, *Prod. Oper. Manag.*, **25** (2016), 1513–1527. <https://doi.org/10.1111/poms.12565>
12. K. Li, J. Wang, Equilibrium balking strategies in the single-server retrial queue with constant retrial rate and catastrophes, *Qual. Technol. Quant. M.*, **18** (2021), 156–178. <https://doi.org/10.1080/16843703.2020.1760464>
13. Z. Meng, L. Liu, X. Chai, Z. Wang, Equilibrium strategies in a constant retrial queue with setup time and the N-policy, *Commun. Stat.-Theory M.*, **49** (2020), 1695–1711. <https://doi.org/10.1080/03610926.2019.1565779>
14. C. Öz, F. Karaesmen, On a production/inventory system with strategic customers and unobservable inventory levels, *SMMSO*, 2015, 161.

15. M. A. Reiyas, K. Jeganathan, A classical retrial queueing inventory system with two component demand rate, *Int. J. Operat. Res.*, **74** (2023), 508–533. <https://doi.org/10.1504/IJOR.2023.132813>
16. A. Krishnamoorthy, D. Shajin, Stochastic decomposition in retrial queueing-inventory system, *RAIRO-Oper. Res.*, **54** (2020), 88–91. <https://doi.org/10.1145/3016032.3016043>
17. X. Shi, L. Liu, Equilibrium joining strategies in the retrial queue with two classes of customers and delayed vacations, *Methodol. Comput. Appl.*, **25** (2023), 1–27. <https://doi.org/10.1007/s11009-023-10029-y>
18. J. Wang, F. Zhang, Strategic joining in m/m/1 retrial queues, *Eur. J. Oper. Res.*, **230** (2013), 76–87. <https://doi.org/10.1016/j.ejor.2013.03.030>
19. Z. Wang, L. Liu, Y. Q. Zhao, Equilibrium customer and socially optimal balking strategies in a constant retrial queue with multiple vacations and N-policy, *J. Comb. Optim.*, **43** (2022), 870–908. <https://doi.org/10.1007/s10878-021-00814-1>
20. J. Wang, X. Zhang, P. Huang, Strategic behavior and social optimization in a constant retrial queue with the N-policy, *Eur. J. Oper. Res.*, **256** (2017), 841–849. <https://doi.org/10.1016/j.ejor.2016.06.034>
21. Y. Zhang, Strategic behavior in the constant retrial queue with a single vacation, *RAIRO-Oper. Res.*, **54** (2020), 569–583. <https://doi.org/10.1051/ro/2019016>
22. X. Zhang, J. Wang, Optimal inventory threshold for a dynamic service make-to-stock system with strategic customers, *Appl. Stoch. Model. Bus.*, **35** (2019), 1103–1123. <https://doi.org/10.1002/asmb.2454>
23. Y. Zhang, J. Wang, Managing retrial queueing systems with boundedly rational customers, *J. Oper. Res. Soc.*, **74** (2023), 748–761. <https://doi.org/10.1080/01605682.2022.2053305>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)