Queueing Models and
Service Management
Vol. 8, No. 1, page 1–20, 2025



# A Batch Demands Queueing-inventory Model with Positive Replenishment Time

Akash Verma[1], Sujit Kumar Samanta[1,*], K.P. Sapna Isotupa[2]

[1]Department of Mathematics
National Institute of Technology Raipur
Raipur 492010, Chhattisgarh, India

[2]School of Business and Economics
Wilfrid Laurier University
Waterloo, Ontario, N2L 3C5, Canada

(*Received April 2023; accepted October 2024*)

**Abstract:** This study examines an $M/M/1$ queueing model using a $(s, Q)$ inventory system and batch demand environment. The customers may request one or more items with a maximum demand of $B$ items. There is just one service provider and a limited number of places available for customers to wait. A customer who arrived in accordance with the Poisson process is forced to leave without service if the waiting space is full. The service time and the replenishment time are assumed to have independent exponential distributions. It is assumed that maximum batch demand $B$ is smaller than or equal to reorder level $s$ in order to avoid repeated replenishment in a service. With the help of an iterative process, we are able to derive the steady-state joint probability distribution of the number of customers in the system and the on-hand inventory level of the queueing-inventory model. The optimum values for waiting space $(N)$, reorder level $(s)$, and order quantity $(Q)$ are found by establishing a number of stationary system performance measures and estimating the total expected cost function under an appropriate cost structure. Some numerical results for various model parameters are provided in order to explain the key performance measures of the system. We execute simulation results with ARENA software to validate our model. We also perform simulation results for the equivalent $M/G/1$ queueing-inventory model.

**Keywords:** Batch demands, cost optimization, positive replenishment time, queueing-inventory model, $(s, Q)$ inventory policy, simulation

## 1. Introduction

The queueing-inventory model is a type of queueing system that includes an inventory component, where resources are stored and utilized to serve customers. In a queueing-inventory model, the inventory level decreases at the service rate instead of the customer

* Corresponding author
Email: sksamanta.maths@nitrr.ac.in

arrival rate if there are customers waiting in the queue to be served. The service will be discontinued if there is no inventory available and/or no customer in the system. Over the years, researchers have received considerable attention to the study of queueing systems with inventory control. Customers who need service enter the service facility. An item from the stock is required to finish the customer service. As soon as a customer receives service, they instantly leave the system which results in a one-for-one reduction in the amount of stock on hand. A third party vendor provides the inventory. We are aware that the inventory is often depleted with a single item in most queueing-inventory models. However, customers might have demand for more than one item known as batch demands. The batch demands in queueing-inventory models have drawn a lot of interest because of their potential use in managing high volume inventory. Batch demands reduce cost associated with the fast flow of large amount commodities. The relation between the queue and the service mechanism with batch demands is a key aspect of the batch demand model.

The concept of batch demands can be found in many different contexts throughout our daily life. There is a rising trend in service oriented firms to improve the standard of their services. The queueing-inventory model with batch demands is commonly driven by grocery sector, as customers frequently place bulk orders for basic commodities or goods. People used to be hesitant to leave their homes when the COVID-19 pandemic was around. People started to purchase commodities in significantly large quantities than normal. Because it helped them to safely evacuate from the corona outbreak. Another instance is that customers may like to order one or more meals when they visit to a restaurant for take away meals. In a business-to-business trading process, a retailer often purchases products in huge quantities from a wholesaler or a franchisee buys goods in bulk from his franchisor. The wholesaler or franchisor deals with batch demands and may use queueing-inventory policy to manage their stocks.

A pioneering contribution to the queueing-inventory domain was made by Sigman and Simchi-Levi [20]. Berman et al. [1] worked for inventory control at service facilities in which depletion of one item from stock in a service provider. Berman and Kim [2] investigated the single demand service facility model, where orders are only placed when the stock level reaches zero with instantaneous order replenishments and no orders are placed while the system is empty. Schwarz and Daduna [17], Schwarz et al. [18] and Saffari et al. [15] all devoted on the subject of service facilities with only one item offering for service. Schwarz et al. [19] worked with joint queueing-inventory systems under continuous review with different inventory management policies. A batch service model for an infinite capacity single server queue associated to a $(s, S)$-type inventory system was developed by Chakravarthy et al. [5]. The batch size depends on specified thresholds and available inventory, where every customer in the batch that needs to be served requires single item for service. The reader is recommended to read Krishnamoorthy et al. [12], Karthikeyan and Sudhesh [10] and Krishnamoorthy et al. [13] for a comprehensive review of literature on single item inventory systems with service facilities. Karthick et al. [9] investigated the $(s, S)$ continuous review inventory system in which customers are classified into two groups: type-1 and type-2. Various performance measures are used to estimate the overall expected cost rate, which is derived from the joint probability distribution of the number of customers in queue and the stock level. Keerthana et al. [11] extended postponed inventory system by incorporating various demand distributions for inter-arrival times. The joint probability distribution of the number of customers in the pool and the inventory level is determined using the matrix geometric method. Samanta et al. [16] worked on a service facility with $(s, Q)$ inventory policy on limited waiting space. They not only determined the system length distributions at random and post-departure epochs, but also identified the optimum waiting space, reorder level, and order quantity that minimizes the overall estimated cost. Chakravarthy and Hayat [4] used the well-known matrix-analytic method to study a two-vendor queueing-inventory system wherein the lead times are exponentially distributed with parameter that varies depending on the vendor, the service times are of phase type, and the demands occur according to a Markovian arrival process. Melikov et al. [14] conducted an analysis of a perishable inventory system with delayed feedback under the $(s, S)$ policy, where the customers can either leave the system with/without purchasing an item depending on the state of the server as well as inventory level. Interested readers are referred to the works of Chakravarthy and Subramanian [7], Divya et al. [8], Yue et al. [21], and queueing-inventory system related references therein.

There have been new interest in batch demands which is different from all of the models based on single commodity demand discussed above. However, fewer research has been carried out to address batch demands. To the best of our knowledge, Yue et al. [22] is the first study to handle the batch demands in $M/M/1$ queueing-inventory system. We identified two recent research articles on batch demands carried out by Chakravarthy [3], and Chakravarthy and Rumyantsev [6]. Yue et al. [22] investigated an infinite waiting space $M/M/1$ queueing-inventory system that incorporated batch demands and lost sales (specifically, partial-lost sales and full-lost sales). The batch demand size for each arrival is truncated geometric distribution, ensuring that the batch size is limited. When the amount of items on hand is smaller than the quantity of requests from an arriving customer, a partial-loss sale occurs, in which case the customer takes everything from the inventory; in a full-loss sale, however, the customer departs the store without purchasing anything from the inventory. They made the assumption that the inventory is managed in accordance with the $(s, S)$ policy which used global searching approach to obtain the optimal reorder level $s$ for the specified $S$. Chakravarthy [3] studied two models dealing with an infinite waiting space $M/M/1$ queueing-inventory system that incorporated batch demands. He assumed that the batch size of demands is random with a finite support and analyzed the models under $(s, S)$ policy to manage the inventory. When a customer arrives and the server is empty or is serving a customer and finds there is no inventory, he assumed that all of the customers are lost. The customer's requirements are satisfied to a minimum by the inventory that is available at the start of service and the desired requirements. He presented the steady-state analysis of the models using the conventional matrix-analytic technique. Chakravarthy and Rumyantsev [6] generalized the queueing-inventory model studied in Chakravarthy [3] by taking into account MAP arrivals and PH-type service times. They used the matrix-analytic method to produce the analytical results for the single server model, whereas ARENA simulation was used to examine the multi-server model.

In this paper, we investigate an $M/M/1$ queueing model with $(s, Q)$ inventory policy and batch demands. The waiting space for customers is finite. The demand is considered to be random with finite support and not go over the reorder level. When the requested demand exceeds the inventory on hand, the customer leaves the system with available inventory. An arriving customer get lost only when the waiting area is full. We calculate the joint steady-state probability distribution for the length of customer in system and the amount of inventory that is currently on hand in the queueing-inventory system. Using a few performance criteria, we investigate a whole framework to calculate the overall expected cost. The optimum waiting capacity $(N)$, reordering threshold $(s)$, and order size $(Q)$ are calculated numerically. The structure of the cost function is extremely complex makes it challenging to determine any unimodality. However, using the numerical data, we have graphically demonstrated that the cost function is convex. Key performance measures of the system are explained with a few numerical results for different model parameters. We used ARENA software to validate our model by performing simulation results. The corresponding $M/G/1$ queueing-inventory model was also simulated and the results are shown.

The structure of this paper is as follows: Section 2 presents the description of the queueing-inventory model. Section 3 explores the steady-state joint distribution of inventory level and number of customers in the system. Section 4 contains some key performance measures and cost function. Numerical results are presented in Section 5. The paper concludes in Section 6.

## 2. Model Description

We consider an $M/M/1$ queueing model with $(s, Q)$ inventory policy and batch demands. Under $(s, Q)$ policy, we consider the service facility with a maximum inventory size of $(Q + s)$ units excluding the items in service. Customers enter the system following a Poisson process with a rate of $\lambda > 0$. The waiting space is $N$ excluding one customer in the service, so that an arriving customer who sees $N$ customers in the queue will be rejected. The service time is assumed to be exponential distribution with a rate of $\mu > 0$. Each arrival may demand one or more items, but they are not permitted to make requests for more than $B$ items. It is assumed that the demands' batch size $X$ is random with probability mass function (p.m.f.) $P(X = i) = \alpha_i$, $i = 1, 2, \ldots, B$. Furthermore, let $\tilde{\alpha}_j = \sum_{k=j}^{B} \alpha_k$, where $j = 1, \ldots, B$. If the requested demand is greater than the available inventory then the customer leaves the system with the available inventory. In order to avoid repeated replenishment in a service, we assume that batch size $B$ is below or equal to reorder threshold $s$. The service will not start until the next replenishment arrives if the server is prepared to serve a customer but there is no inventory available. When the inventory level hits a predefined value $s$, the quantity $Q$, $(Q > s)$ is ordered. An external supplier replenishes the inventory using an exponential distribution with a rate of $\gamma > 0$. Figure 1 illustrates the proposed batch demand model.

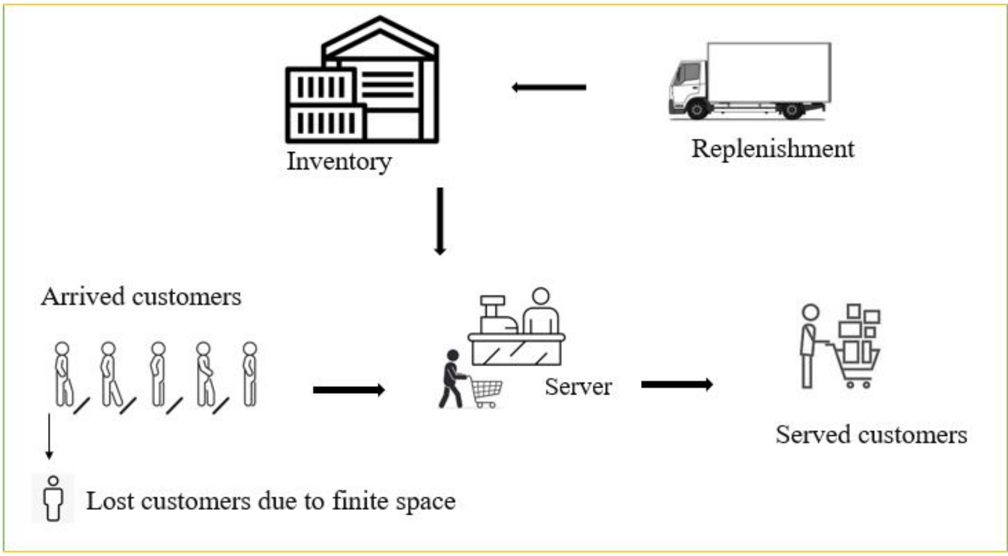


Figure 1. Flowchart for the proposed batch demand model.

## 3. Steady State Analysis

In this section, we determine the steady-state joint probability distribution of inventory level in stock and number of customers in the system. For this purpose, we define the following random variables at time $t$ to establish the state of the system:

- $I(t)$ = number of items present in inventory (stock) excluding items in service (if any),
- $N(t)$ = number of customers present in the system,
- $\xi(t)$ = state of the server, $\xi(t) = \begin{cases} 0, & \text{server is idle,} \\ 1, & \text{server is busy.} \end{cases}$

The state of the system at time $t$ is defined by the Markov chain $\{I(t), N(t), \xi(t) : t \geq 0\}$. In the steady-state, let us define their joint probabilities as

$$\omega(i,0) = \lim_{t\to\infty} P[I(t) = i, N(t) = 0, \xi(t) = 0], \quad i = 0,1,2,\ldots,Q+s,$$
$$\upsilon(0,n) = \lim_{t\to\infty} P[I(t) = 0, N(t) = n, \xi(t) = 0], \quad n = 1,\ldots,N,$$
$$\pi(i,n) = \lim_{t\to\infty} P[I(t) = i, N(t) = n, \xi(t) = 1], \quad i = 0,1,2,\ldots,Q+s, \quad n = 1,\ldots,N+1.$$

We first construct the following difference equations in the steady-state by joining the states of the system at time $t$ and $t + dt$ in order to determine the steady-state probabilities defined above. Using the probabilistic argument, we considered the following five scenarios separately based on the number of customers in the system and the current inventory level.

**Case 1: When the server is busy and the queue is empty**

$$(\lambda + \mu)\pi(Q+s,1) = \gamma\pi(s,1), \tag{1}$$

$$(\lambda+\mu)\pi(l,1)=\mu\sum_{k=1}^{Q+s-l}\alpha_k\pi(l+k,2)+\lambda\sum_{k=1}^{Q+s-l}\alpha_k\omega(l+k,0)+\gamma\pi(l-Q,1),$$
$$Q+s-B\le l\le Q+s-1, \tag{2}$$

$$(\lambda+\mu)\pi(l,1)=\mu\sum_{k=1}^{B}\alpha_k\pi(l+k,2)+\lambda\sum_{k=1}^{B}\alpha_k\omega(l+k,0)+\gamma\pi(l-Q,1),$$
$$Q\le l\le Q+s-B-1, \tag{3}$$

$$(\lambda+\mu)\pi(l,1)=\mu\sum_{k=1}^{B}\alpha_k\pi(l+k,2)+\lambda\sum_{k=1}^{B}\alpha_k\omega(l+k,0)+\gamma\alpha_{Q-l}\upsilon(0,1),$$
$$Q-B\le l\le Q-1, \tag{4}$$

$$(\lambda+\mu)\pi(l,1)=\mu\sum_{k=1}^{B}\alpha_k\pi(l+k,2)+\lambda\sum_{k=1}^{B}\alpha_k\omega(l+k,0),$$
$$s+1\le l\le Q-B-1, \tag{5}$$

$$(\lambda+\mu+\gamma)\pi(l,1)=\mu\sum_{k=1}^{B}\alpha_k\pi(l+k,2)+\lambda\sum_{k=1}^{B}\alpha_k\omega(l+k,0),\quad 1\le l\le s, \tag{6}$$

$$(\lambda+\mu+\gamma)\pi(0,1)=\mu\sum_{k=1}^{B}\tilde{\alpha}_k\pi(k,2)+\lambda\sum_{k=1}^{B}\tilde{\alpha}_k\omega(k,0). \tag{7}$$

**Case 2: When the server is busy and the queue is not full**

$$(\lambda+\mu)\pi(Q+s,n)=\lambda\pi(Q+s,n-1)+\gamma\pi(s,n),\quad 2\le n\le N, \tag{8}$$

$$(\lambda+\mu)\pi(l,n)=\mu\sum_{k=1}^{Q+s-l}\alpha_k\pi(l+k,n+1)+\lambda\pi(l,n-1)+\gamma\pi(l-Q,n),$$
$$Q+s-B\le l\le Q+s-1,\ 2\le n\le N, \tag{9}$$

$$(\lambda+\mu)\pi(l,n)=\mu\sum_{k=1}^{B}\alpha_k\pi(l+k,n+1)+\lambda\pi(l,n-1)+\gamma\pi(l-Q,n),$$
$$Q\le l\le Q+s-B-1,\ 2\le n\le N, \tag{10}$$

$$(\lambda+\mu)\pi(l,n)=\mu\sum_{k=1}^{B}\alpha_k\pi(l+k,n+1)+\lambda\pi(l,n-1)+\gamma\alpha_{Q-l}\upsilon(0,n),$$
$$Q-B\le l\le Q-1,\ 2\le n\le N, \tag{11}$$

$$(\lambda+\mu)\pi(l,n)=\mu\sum_{k=1}^{B}\alpha_k\pi(l+k,n+1)+\lambda\pi(l,n-1),$$
$$s+1\le l\le Q-B-1,\ 2\le n\le N, \tag{12}$$

$$(\lambda+\mu+\gamma)\pi(l,n) = \mu\sum_{k=1}^{B}\alpha_k\pi(l+k,n+1) + \lambda\pi(l,n-1),$$
$$1 \le l \le s,\ 2 \le n \le N, \tag{13}$$
$$(\lambda+\mu+\gamma)\pi(0,n) = \mu\sum_{k=1}^{B}\tilde{\alpha}_k\pi(k,n+1) + \lambda\pi(0,n-1), \quad 2 \le n \le N. \tag{14}$$

**Case 3: When the server is busy and the queue is full**

$$\mu\pi(l,N+1) = \lambda\pi(l,N) + \gamma\pi(l-Q,N+1), \quad Q \le l \le Q+s, \tag{15}$$
$$\mu\pi(l,N+1) = \lambda\pi(l,N), \quad s+1 \le l \le Q-1, \tag{16}$$
$$(\mu+\gamma)\pi(l,N+1) = \lambda\pi(l,N), \quad 0 \le l \le s. \tag{17}$$

**Case 4: When the server is idle and no customers in the system**

$$\lambda\omega(l,0) = \mu\pi(l,1) + \gamma\omega(l-Q,0), \quad Q \le l \le Q+s, \tag{18}$$
$$\lambda\omega(l,0) = \mu\pi(l,1), \quad s+1 \le l \le Q-1, \tag{19}$$
$$(\lambda+\gamma)\omega(l,0) = \mu\pi(l,1), \quad 0 \le l \le s. \tag{20}$$

**Case 5: When the server is idle and no items in inventory**

$$\gamma\upsilon(0,N) = \lambda\upsilon(0,N-1) + \mu\pi(0,N+1), \tag{21}$$
$$(\lambda+\gamma)\upsilon(0,n) = \lambda\upsilon(0,n-1) + \mu\pi(0,n+1), \quad 2 \le n \le N-1, \tag{22}$$
$$(\lambda+\gamma)\upsilon(0,1) = \lambda\omega(0,0) + \mu\pi(0,2). \tag{23}$$

Now, we use the following iterative steps to express $\upsilon(0,n)$, $n = 1,2,\ldots,N$; $\omega(l,0)$, $l = 0,1,\ldots,Q+s$ and $\pi(l,n)$, $l = s+1, s+2, \ldots, Q+s$, $n = 1,2,\ldots,N+1$ in terms of $\pi(l,n)$, $l = 0,1,2,\ldots,s$, $n = 1,2,\ldots,N+1$ in order to evaluate all the steady-state probabilities. From (20), (23), (22) and (21), respectively, we obtain

$$\omega(l,0) = \left(\frac{\mu}{\lambda+\gamma}\right)\pi(l,1), \quad 0 \le l \le s,$$
$$\upsilon(0,1) = \left(\frac{\lambda}{\lambda+\gamma}\right)\omega(0,0) + \left(\frac{\mu}{\lambda+\gamma}\right)\pi(0,2),$$
$$\upsilon(0,n) = \left(\frac{\lambda}{\lambda+\gamma}\right)\upsilon(0,n-1) + \left(\frac{\mu}{\lambda+\gamma}\right)\pi(0,n+1), \quad n = 2,3,\ldots,N-1,$$
$$\upsilon(0,N) = \left(\frac{\lambda}{\gamma}\right)\upsilon(0,N-1) + \left(\frac{\mu}{\gamma}\right)\pi(0,N+1).$$

From (1), (18), (8) and (15), respectively, we obtain

$$\pi(Q+s,1) = \left(\frac{\gamma}{\lambda+\mu}\right)\pi(s,1),$$

$$\omega(Q+s,0) = \left(\frac{\mu}{\lambda}\right)\pi(Q+s,1) + \left(\frac{\gamma}{\lambda}\right)\omega(s,0),$$
$$\pi(Q+s,n) = \left(\frac{\lambda}{\lambda+\mu}\right)\pi(Q+s,n-1) + \left(\frac{\gamma}{\lambda+\mu}\right)\pi(s,n), \quad n = 2,3,\ldots,N,$$
$$\pi(Q+s,N+1) = \left(\frac{\lambda}{\mu}\right)\pi(Q+s,N) + \left(\frac{\gamma}{\mu}\right)\pi(s,N+1).$$

Now, $\pi(l,1)$, $\omega(l,0)$ and $\pi(l,n)$, for $2 \le n \le N$; $Q+s-B \le l \le Q+s-1$, are obtained from (2), (18) and (9), respectively, with the following iterative procedure:

$$\pi(l,1) = \left(\frac{\mu}{\lambda+\mu}\right)\sum_{k=1}^{Q+s-l}\alpha_k\pi(l+k,2) + \left(\frac{\lambda}{\lambda+\mu}\right)\sum_{k=1}^{Q+s-l}\alpha_k\omega(l+k,0) + \left(\frac{\gamma}{\lambda+\mu}\right)\pi(l-Q,1),$$
$$l = Q+s-1, Q+s-2, \ldots, Q+s-B,$$
$$\omega(l,0) = \left(\frac{\mu}{\lambda}\right)\pi(l,1) + \left(\frac{\gamma}{\lambda}\right)\omega(l-Q,0), \quad l = Q+s-1, Q+s-2, \ldots, Q+s-B,$$
$$\pi(l,n) = \left(\frac{\mu}{\lambda+\mu}\right)\sum_{k=1}^{Q+s-l}\alpha_k\pi(l+k,n+1) + \left(\frac{\lambda}{\lambda+\mu}\right)\pi(l,n-1) + \left(\frac{\gamma}{\lambda+\mu}\right)\pi(l-Q,n),$$
$$l = Q+s-1, Q+s-2, \ldots, Q+s-B,\ n = 2,3,\ldots,N.$$

Similarly, $\pi(l,1)$, $\omega(l,0)$ and $\pi(l,n)$, for $2 \le n \le N$; $Q \le l \le Q+s-B-1$, are obtained from (3), (18) and (10), respectively, with the following iterative procedure:

$$\pi(l,1) = \left(\frac{\mu}{\lambda+\mu}\right)\sum_{k=1}^{B}\alpha_k\pi(l+k,2) + \left(\frac{\lambda}{\lambda+\mu}\right)\sum_{k=1}^{B}\alpha_k\omega(l+k,0) + \left(\frac{\gamma}{\lambda+\mu}\right)\pi(l-Q,1),$$
$$l = Q+s-B-1, Q+s-B-2, \ldots, Q,$$
$$\omega(l,0) = \left(\frac{\mu}{\lambda}\right)\pi(l,1) + \left(\frac{\gamma}{\lambda}\right)\omega(l-Q,0), \quad l = Q+s-B-1, Q+s-B-2, \ldots, Q,$$
$$\pi(l,n) = \left(\frac{\mu}{\lambda+\mu}\right)\sum_{k=1}^{B}\alpha_k\pi(l+k,n+1) + \left(\frac{\lambda}{\lambda+\mu}\right)\pi(l,n-1) + \left(\frac{\gamma}{\lambda+\mu}\right)\pi(l-Q,n),$$
$$l = Q+s-B-1, Q+s-B-2, \ldots, Q,\ n = 2,3,\ldots,N.$$

From (15), we repeatedly obtain as

$$\pi(l,N+1) = \left(\frac{\lambda}{\mu}\right)\pi(l,N) + \left(\frac{\gamma}{\mu}\right)\pi(l-Q,N+1), \quad l = Q+s-1, Q+s-2, \ldots, Q.$$

Further, $\pi(l,1)$, $\omega(l,0)$ and $\pi(l,n)$, for $2 \le n \le N$; $Q-B \le l \le Q-1$, are obtained from (4), (19) and (11), respectively, with the following iterative procedure:

$$\pi(l,1) = \left(\frac{\mu}{\lambda+\mu}\right)\sum_{k=1}^{B}\alpha_k\pi(l+k,2) + \left(\frac{\lambda}{\lambda+\mu}\right)\sum_{k=1}^{B}\alpha_k\omega(l+k,0) + \left(\frac{\gamma}{\lambda+\mu}\right)\alpha_{Q-l}\upsilon(0,1),$$

$$
\begin{aligned}
& \qquad\qquad\qquad\qquad\qquad\qquad l = Q-1, Q-2, \ldots, Q-B, \\
\omega(l,0) &= \left(\frac{\mu}{\lambda}\right)\pi(l,1), \quad l = Q-1, Q-2, \ldots, Q-B, \\
\pi(l,n) &= \left(\frac{\mu}{\lambda+\mu}\right)\sum_{k=1}^{B}\alpha_k\pi(l+k,n+1) + \left(\frac{\lambda}{\lambda+\mu}\right)\pi(l,n-1) + \left(\frac{\gamma}{\lambda+\mu}\right)\alpha_{Q-l}\upsilon(0,n), \\
& \qquad\qquad l = Q-1, Q-2, \ldots, Q-B, \; n = 2,3,\ldots,N.
\end{aligned}
$$

From (16), we repeatedly obtain as

$$
\pi(l,N+1) = \left(\frac{\lambda}{\mu}\right)\pi(l,N), \quad l = Q-1, Q-2, \ldots, Q-B.
$$

Similarly, $\pi(l,1)$, $\omega(l,0)$ and $\pi(l,n)$, for $2 \le n \le N$; $s+1 \le l \le Q-B-1$, are obtained from (5), (19) and (12), respectively, with the following iterative procedure:

$$
\begin{aligned}
\pi(l,1) &= \left(\frac{\mu}{\lambda+\mu}\right)\sum_{k=1}^{B}\alpha_k\pi(l+k,2) + \left(\frac{\lambda}{\lambda+\mu}\right)\sum_{k=1}^{B}\alpha_k\omega(l+k,0), \\
& l = Q-B-1, Q-B-2, \ldots, s+1, \\
\omega(l,0) &= \left(\frac{\mu}{\lambda}\right)\pi(l,1), \quad l = Q-B-1, Q-B-2, \ldots, s+1, \\
\pi(l,n) &= \left(\frac{\mu}{\lambda+\mu}\right)\sum_{k=1}^{B}\alpha_k\pi(l+k,n+1) + \left(\frac{\lambda}{\lambda+\mu}\right)\pi(l,n-1), \\
& l = Q-B-1, Q-B-2, \ldots, s+1, \; n = 2,3,\ldots,N.
\end{aligned}
$$

From (16), we repeatedly obtain as

$$
\pi(l,N+1) = \left(\frac{\lambda}{\mu}\right)\pi(l,N), \quad l = Q-B-1, Q-B-2, \ldots, s+1.
$$

Now, the $(N+1)(s+1)$ system of simultaneous linear equations in $(N+1)(s+1)$ unknowns are obtained from (6), (7), (13), (14), and (17) using the above results. Solving these $(N+1)(s+1)$ system of simultaneous linear equations by replacing any one equation with the normalization equation $\sum_{n=1}^{N}\upsilon(0,n)+\sum_{l=0}^{Q+s}\omega(l,0)+\sum_{l=0}^{Q+s}\sum_{n=1}^{N+1}\pi(l,n)=1$, we get the solution of $\pi(l,n)$, $l = 0,1,2,\ldots,s$, $n = 1,2,\ldots,N+1$. The above iterative procedures can be used to obtain the remaining probabilities $\upsilon(0,n)$, $n = 1,2,\ldots,N$; $\omega(l,0)$, $l = 0,1,\ldots,Q+s$ and $\pi(l,n)$, $l = s+1, s+2, \ldots, Q+s$, $n = 1,2,\ldots,N+1$.

## 4. Some Performance Measures with Cost Function

Here, we use the steady-state probabilities derived in Section 3 to obtain the following performance measures and use them to develop the cost function.

The average number of inventory $(I_v)$ in stock is given by

$$I_v = \sum_{l=1}^{Q+s} l\omega(l,0) + \sum_{l=1}^{Q+s} l \sum_{j=1}^{N+1} \pi(l,j).$$

The average number of customers in the queue $(I_c)$ is expressed by

$$I_c = \sum_{j=1}^{N} j\upsilon(0,j) + \sum_{j=1}^{N+1} (j-1) \sum_{l=0}^{Q+s} \pi(l,j).$$

Let $R_o$ represent the mean replenishment rate. Then, it can be determined as

$$R_o = \gamma \left( \sum_{j=1}^{N} \upsilon(0,j) + \sum_{l=0}^{s} \omega(l,0) + \sum_{l=0}^{s} \sum_{j=1}^{N+1} \pi(l,j) \right).$$

Let $R_c$ be the rate of rejection of a customer from the system. Then, it can be represented as

$$R_c = \lambda \left( \upsilon(0,N) + \sum_{l=0}^{Q+s} \pi(l,N+1) \right).$$

The cost function $C(N,s,Q)$ per unit time for the proposed model is given as

$$C(N,s,Q) = h_c I_c + h_v I_v + (s_c + p_c Q) R_o + r_c R_c,$$

where the cost coefficients are as follows:

- $h_c$ (holding cost) is the cost per customer per unit time,
- $h_v$ (inventory carrying cost) is the cost per item per unit time,
- $s_c$ (setup cost) is the cost per order,
- $p_c$ (purchase price) is the cost per item,
- $r_c$ (rejection cost) is the cost of rejection of a customer because of full system capacity.

## 5. Numerical Analysis

This section devotes to observe several interesting behaviours in the analysis of system performance measures and cost function with various parameter settings. We provide performance measures and optimum results in the forms of tables and graphs for three different batch sizes such as $B = 4$, $B = 6$, and $B = 8$. We use $\alpha_1 = 0.35$, $\alpha_2 = 0.15$, $\alpha_3 = 0.25$, $\alpha_4 = 0.25$, when $B = 4$. We use $\alpha_1 = 0.25$, $\alpha_2 = 0.15$, $\alpha_3 = 0.15$, $\alpha_4 = 0.30$, $\alpha_5 = 0.05$, $\alpha_6 = 0.10$, when $B = 6$. We use $\alpha_1 = 0.25$, $\alpha_2 = 0.15$, $\alpha_3 = 0.05$, $\alpha_4 = 0.20$, $\alpha_5 = 0.05$, $\alpha_6 = 0.10$, $\alpha_7 = 0.05$, $\alpha_8 = 0.15$, when $B = 8$. The impact of arrival rate $\lambda$ on various performance measures is shown in Table 1. The table shows that the values of $I_v$ seem to decrease when $\lambda$ increases, whereas $I_c$, $R_o$, and $R_c$ show a rising tendency as $\lambda$ values rise for each batch demand size. This implies that when $\lambda$ increases, the average number of customers waiting in the queue also increases and this leads to more rejections because of waiting space is fixed. The effect of service rate $\mu$ on performance measures is shown in Table 2, where the values of $I_v$, $I_c$, and $R_c$ seem to decrease with an increase of $\mu$, while an increasing trend is noted with higher values of $\mu$ in the case of $R_o$. As the service rate increases, the average number of customers in the queue decreases as more services are performed. Rejections are reduced as a result of the fixed waiting space. Table 3 demonstrates the effect of the replenishment rate $\gamma$. The table illustrates that the values of $I_v$ and $R_o$ seem to increase as the value of $\gamma$ rises, whereas a downward tendency is observed with higher values of $\gamma$ in the case of $I_c$. Furthermore, $R_c$ remain almost constant as $\gamma$ increases. The lower mean replenishment rate causes an increase in inventory levels. The impact of $N$ is shown in Table 4. The table shows that the values of $I_v$ and $R_c$ tend to decrease with rising $N$, while the values of $I_c$ and $R_o$ show an increasing trend with higher values of $N$. There are more waiting spaces when $N$ rises, which raises the average number of customers in the queue and reduces the number of rejections. Table 5 shows the effect of the ordering point $s$ on various performance measures. The information suggests that the values of $I_v$ and $R_o$ tend to increase as $s$ rises. On the other hand, the values of $I_c$ decreases with higher values of $s$ whereas $R_c$ decreases relatively very little as $s$ increases. This is because of frequent replenishment when $s$ is higher for fix $Q$. Moreover, Table 6 demonstrates the effect of $Q$. The data indicates that the value of $I_v$ tends to increase as $Q$ rises, while the values of $I_c$ and $R_o$ tend to decrease as $Q$ rises. Further, $R_c$ decreases relatively very little as $Q$ increases. The larger order quantities reduce the frequency of orders required, which raises the average inventory level.

Table 1. Effect of $\lambda$: Fix $Q = 25$, $s = 12$, $N = 14$, $\mu = 6$, $\gamma = 3$

| | **B=4** | | | | **B=6** | | | | **B=8** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | 4 | 5 | 6 | 7 | 4 | 5 | 6 | 7 | 4 | 5 | 6 | 7 |
| $I_v$ | 21.82 | 21.08 | 20.56 | 20.37 | 20.98 | 20.06 | 19.45 | 19.24 | 19.93 | 18.80 | 18.13 | 17.93 |
| $I_c$ | 1.36 | 3.46 | 6.92 | 9.80 | 1.43 | 3.69 | 7.26 | 10.04 | 1.58 | 4.13 | 7.82 | 10.42 |
| $R_o$ | 0.38 | 0.47 | 0.53 | 0.56 | 0.48 | 0.59 | 0.67 | 0.69 | 0.61 | 0.75 | 0.83 | 0.85 |
| $R_c$ | 0.00 | 0.07 | 0.42 | 1.18 | 0.00 | 0.08 | 0.47 | 1.26 | 0.01 | 0.10 | 0.57 | 1.40 |

Table 2. Effect of $\mu$: Fix $Q = 25$, $s = 12$, $N = 14$, $\lambda = 6$, $\gamma = 3$

| | **B=4** | | | | **B=6** | | | | **B=8** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mu$ | 4 | 5 | 7 | 8 | 4 | 5 | 7 | 8 | 4 | 5 | 7 | 8 |
| $I_v$ | 21.83 | 21.10 | 20.33 | 20.27 | 21.02 | 20.12 | 19.14 | 19.05 | 20.03 | 18.96 | 17.70 | 17.55 |
| $I_c$ | 12.07 | 10.10 | 4.05 | 2.38 | 12.11 | 10.26 | 4.42 | 2.68 | 12.19 | 10.53 | 5.06 | 3.23 |
| $R_o$ | 0.38 | 0.47 | 0.56 | 0.57 | 0.48 | 0.59 | 0.71 | 0.72 | 0.60 | 0.73 | 0.88 | 0.90 |
| $R_c$ | 2.02 | 1.10 | 0.12 | 0.03 | 2.05 | 1.15 | 0.15 | 0.05 | 2.10 | 1.24 | 0.21 | 0.08 |

Table 3. Effect of $\gamma$: Fix $Q = 25$, $s = 12$, $N = 14$, $\lambda = 4$, $\mu = 6$

| | **B=4** | | | | **B=6** | | | | **B=8** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\gamma$ | 4 | 5 | 6 | 7 | 4 | 5 | 6 | 7 | 4 | 5 | 6 | 7 |
| $I_v$ | 22.61 | 23.08 | 23.40 | 23.63 | 21.97 | 22.57 | 22.97 | 23.26 | 21.17 | 21.92 | 22.43 | 22.79 |
| $I_c$ | 1.33 | 1.32 | 1.31 | 1.31 | 1.36 | 1.33 | 1.32 | 1.32 | 1.42 | 1.37 | 1.34 | 1.33 |
| $R_o$ | 0.38 | 0.38 | 0.38 | 0.38 | 0.49 | 0.49 | 0.49 | 0.49 | 0.62 | 0.62 | 0.62 | 0.62 |
| $R_c$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Table 4. Effect of $N$: Fix $Q = 25$, $s = 12$, $\lambda = 4$, $\mu = 6$, $\gamma = 3$

| | **B=4** | | | | **B=6** | | | | **B=8** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | 9 | 10 | 11 | 12 | 9 | 10 | 11 | 12 | 9 | 10 | 11 | 12 |
| $I_v$ | 21.84 | 21.83 | 21.82 | 21.82 | 21.00 | 20.99 | 20.99 | 20.98 | 19.96 | 19.95 | 19.94 | 19.93 |
| $I_c$ | 1.25 | 1.29 | 1.31 | 1.34 | 1.30 | 1.34 | 1.38 | 1.40 | 1.41 | 1.46 | 1.50 | 1.54 |
| $R_o$ | 0.38 | 0.38 | 0.38 | 0.38 | 0.48 | 0.48 | 0.48 | 0.48 | 0.61 | 0.61 | 0.61 | 0.61 |
| $R_c$ | 0.03 | 0.02 | 0.01 | 0.01 | 0.03 | 0.02 | 0.01 | 0.01 | 0.04 | 0.03 | 0.02 | 0.01 |

Table 5. Effect of $s$: Fix $Q = 25$, $N = 14$, $\lambda = 4$, $\mu = 6$, $\gamma = 3$

| | **B=4** | | | | **B=6** | | | | **B=8** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $s$ | 9 | 10 | 11 | 12 | 9 | 10 | 11 | 12 | 9 | 10 | 11 | 12 |
| $I_v$ | 18.87 | 19.85 | 20.83 | 21.82 | 18.10 | 19.05 | 20.01 | 20.98 | 17.17 | 18.07 | 19.00 | 19.93 |
| $I_c$ | 1.41 | 1.39 | 1.38 | 1.36 | 1.52 | 1.49 | 1.46 | 1.43 | 1.72 | 1.67 | 1.62 | 1.58 |
| $R_o$ | 0.38 | 0.38 | 0.38 | 0.38 | 0.48 | 0.48 | 0.48 | 0.48 | 0.61 | 0.61 | 0.61 | 0.61 |
| $R_c$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 |

Table 6. Effect of $Q$: Fix $s = 12$, $N = 14$, $\lambda = 4$, $\mu = 6$, $\gamma = 3$

| | **B=4** | | | | **B=6** | | | | **B=8** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Q$ | 21 | 22 | 23 | 24 | 21 | 22 | 23 | 24 | 21 | 22 | 23 | 24 |
| $I_v$ | 19.80 | 20.30 | 20.81 | 21.32 | 18.94 | 19.45 | 19.96 | 20.47 | 17.82 | 18.35 | 18.88 | 19.42 |
| $I_c$ | 1.37 | 1.37 | 1.37 | 1.37 | 1.46 | 1.45 | 1.45 | 1.44 | 1.64 | 1.63 | 1.61 | 1.59 |
| $R_o$ | 0.46 | 0.43 | 0.41 | 0.40 | 0.58 | 0.55 | 0.53 | 0.50 | 0.73 | 0.69 | 0.66 | 0.64 |
| $R_c$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 |

We now examine the optimum values of waiting capacity ($N^*$), reorder point ($s^*$) and order size ($Q^*$) that minimize overall costs. It is difficult to establish the unimodality of the cost function due to its incredibly complex structure. The entire cost value is impacted by a few cost factors which are studied in previous sections. We carry out comprehensive numerical experiments to verify the efficiency of our approach and test for unimodality. It is not necessary to guarantee that the $\rho = \lambda/\mu$ (utilisation factor) should be smaller than 1 because there is a limited number of waiting space for customers. However, we employ a straightforward numerical search technique for parameters to identify the optimum values of $N$, $s$ and $Q$. To determine the optimum results for waiting capacity, reorder point, and order size at minimum cost, we use different rates for arrival of customers and their services. For this numerical analysis, we assume the cost parameters as $h_c = 2$, $h_v = 10$, $s_c = 100$, $p_c = 10$ and $r_c = 200$.

The results for the case where $\lambda$ varies but the mean replenishment is 0.5 and the service rate is $\mu = 6.0$ customers/unit time are shown in Table 7. The arrival rate $\lambda$ varies between

3.4 customers per unit time to 9.0 customers per unit time. It is observed in Table 7 that the optimal waiting capacity decreases as $\rho$ increases. This suggests that we should offer smaller waiting area when service rate is less than arrival rate. We also observe that the optimal order quantity increases when the arrival rate increases and $\rho \leq 1$, but the order quantity remains unchanged when $\rho > 1$. This implies that we should order a specific quantity in order to manage heavy traffic. For $B = 4$ and $B = 6$, we also find that the optimal reorder level increases marginally when the arrival rate rises and $\rho \leq 1$, but the reorder level remains unchanged when $\rho > 1$. The reorder level is not impacted by $\lambda$ for $B = 8$. This implies that we should place an order when inventory level hits the maximum level of demand. Our findings indicate that as $\rho$ increases, so does the optimal cost. This shows that the cost goes up in situations with heavy traffic.

Table 7. The effect of arrival rate $\lambda$ on $C(N, s, Q)$

| | | **B=4** | **B=6** | **B=8** |
|---|---|---|---|---|
| | $\rho = \frac{\lambda}{\mu}$ | $\mu = 6.0, \gamma = 2.0$ | $\mu = 6.0, \gamma = 2.0$ | $\mu = 6.0, \gamma = 2.0$ |
| $\lambda = 3.4$ | 0.57 | $C(57, 4, 12) = 216.08$ | $C(54, 6, 13) = 260.01$ | $C(46, 8, 17) = 310.72$ |
| $\lambda = 4.6$ | 0.77 | $C(47, 4, 17) = 273.68$ | $C(44, 6, 19) = 326.35$ | $C(37, 8, 21) = 388.67$ |
| $\lambda = 5.5$ | 0.92 | $C(26, 6, 23) = 382.25$ | $C(23, 6, 26) = 444.04$ | $C(20, 8, 28) = 517.93$ |
| $\lambda = 6.0$ | 1.0 | $C(20, 7, 24) = 466.11$ | $C(18, 7, 27) = 531.48$ | $C(16, 8, 29) = 607.66$ |
| $\lambda = 7.0$ | 1.17 | $C(13, 8, 24) = 652.06$ | $C(12, 8, 27) = 719.66$ | $C(11, 8, 30) = 797.90$ |
| $\lambda = 7.5$ | 1.25 | $C(11, 8, 24) = 748.49$ | $C(11, 8, 27) = 816.60$ | $C(10, 8, 30) = 895.24$ |
| $\lambda = 8.0$ | 1.33 | $C(10, 8, 24) = 845.92$ | $C(9, 8, 27) = 914.31$ | $C(9, 8, 30) = 993.25$ |
| $\lambda = 8.5$ | 1.42 | $C(9, 8, 24) = 943.99$ | $C(9, 8, 27) = 1012.61$ | $C(8, 8, 30) = 1091.68$ |
| $\lambda = 9.0$ | 1.5 | $C(8, 8, 24) = 1042.51$ | $C(8, 8, 27) = 1111.15$ | $C(7, 8, 30) = 1190.53$ |

Table 8. The effect of service rate $\mu$ on $C(N, s, Q)$

| | | **B=4** | **B=6** | **B=8** |
|---|---|---|---|---|
| | $\rho = \frac{\lambda}{\mu}$ | $\lambda = 6.0, \gamma = 2.0$ | $\lambda = 6.0, \gamma = 2.0$ | $\lambda = 6.0, \gamma = 2.0$ |
| $\mu = 8.0$ | 0.75 | $C(54, 4, 22) = 328.19$ | $C(50, 6, 25) = 391.68$ | $C(44, 8, 29) = 469.24$ |
| $\mu = 7.5$ | 0.80 | $C(44, 4, 25) = 346.31$ | $C(40, 6, 27) = 411.46$ | $C(35, 8, 31) = 491.51$ |
| $\mu = 7.0$ | 0.86 | $C(35, 6, 24) = 374.74$ | $C(32, 7, 28) = 441.96$ | $C(27, 8, 32) = 522.86$ |
| $\mu = 6.0$ | 1.0 | $C(20, 7, 24) = 466.11$ | $C(18, 7, 27) = 531.48$ | $C(16, 8, 29) = 607.66$ |
| $\mu = 5.5$ | 1.09 | $C(15, 7, 22) = 524.30$ | $C(13, 6, 26) = 585.23$ | $C(13, 8, 27) = 656.76$ |
| $\mu = 4.6$ | 1.30 | $C(9, 5, 20) = 636.27$ | $C(9, 6, 22) = 688.30$ | $C(9, 8, 23) = 751.55$ |
| $\mu = 3.4$ | 1.76 | $C(6, 4, 16) = 791.87$ | $C(6, 6, 17) = 835.90$ | $C(5, 8, 17) = 887.89$ |
| $\mu = 2.5$ | 2.4 | $C(4, 4, 12) = 914.59$ | $C(4, 6, 13) = 953.38$ | $C(4, 8, 17) = 1000.38$ |
| $\mu = 2.3$ | 2.60 | $C(4, 4, 11) = 942.87$ | $C(4, 6, 13) = 980.82$ | $C(4, 8, 17) = 1027.48$ |

The results for the case where $\mu$ varies but the mean replenishment is 0.5 and the arrival rate is $\lambda = 6.0$ customers/unit time are shown in Table 8. The service rate $\mu$ varies between 2.3 customers per unit time to 8.0 customers per unit time. It is observed in Table 8 that the optimal waiting capacity decreases as $\rho$ increases. This suggests that we should offer smaller waiting area when service rate is less than arrival rate. It is also noted that when the service rate increases, the optimum order quantity and reorder level also increase; however, as $\mu$

grows, both parameters begin to decrease. This suggests that the quantity of our order and ordering point should be determined by the service rate. The reorder level is not impacted by $\mu$ when $B = 8$. This implies that we should place an order when inventory level hits the maximum level of demand. Our findings indicate that as $\rho$ increases, so does the optimal cost. Figures 2 to 4 confirm the convex behaviour of the cost function.

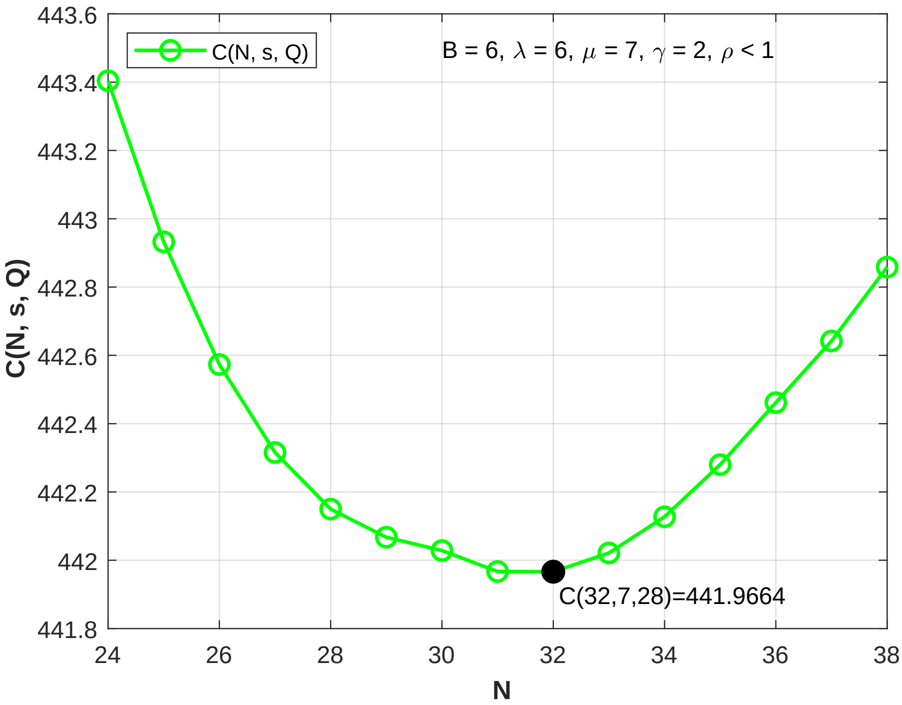


Figure 2. The convex representation of the cost value when $\rho < 1$

We also demonstrate the robustness of our optimal solutions through simulation experiments using the ARENA software. The simulation findings for the $M/M/1$ case, as indicated in Tables 9 to 11 when $\lambda$ varies, have been used to validate the analytical optimal results. It is interesting to see that the outcomes from the two approaches show a similar trend. The validation results indicate relative errors below $1\%$. We simulate the batch demand analysis in the $M/G/1$ case due to its complex nature. The simulation-based comparison results for the $M/M/1$ case versus the $M/G/1$ type scenario are shown in Tables 9 to 11, where non-exponential service time is assumed to be deterministic and normal distributions. We have seen that the results show a similar pattern to what the analytical method indicated for the $M/M/1$ case. Furthermore, the cost behaviour of the simulation results for the $M/G/1$ type and the analytical results for the $M/M/1$ case are shown in Figures 5 to 7 when $\lambda$ varies. Furthermore, as shown by analytical results, we experienced similar patterns during the execution time of the simulation results for the $\mu$ data sets.

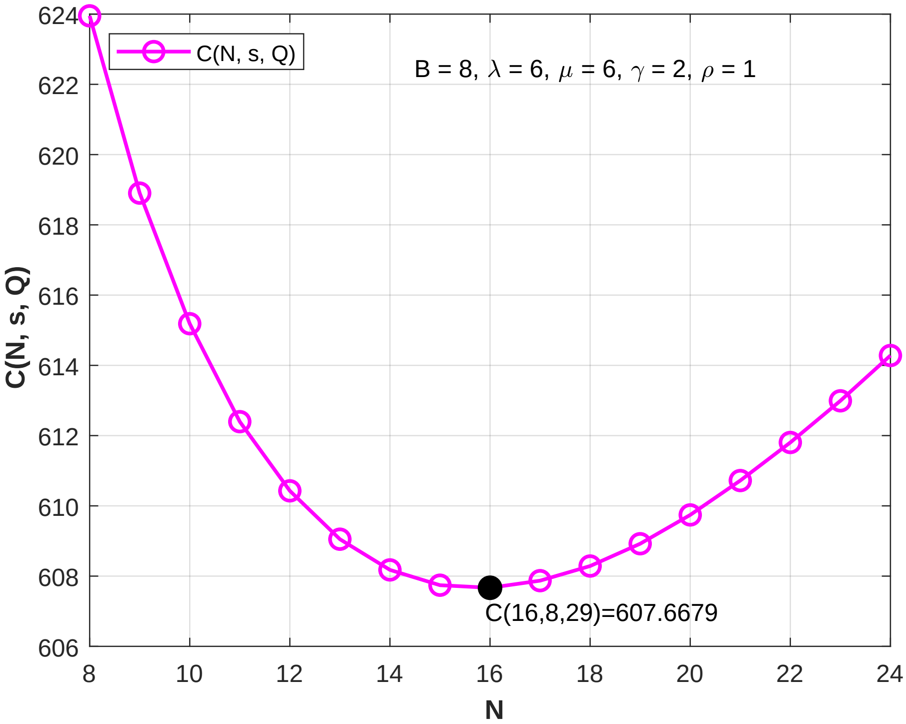


Figure 3. The convex representation of the cost value when $\rho = 1$

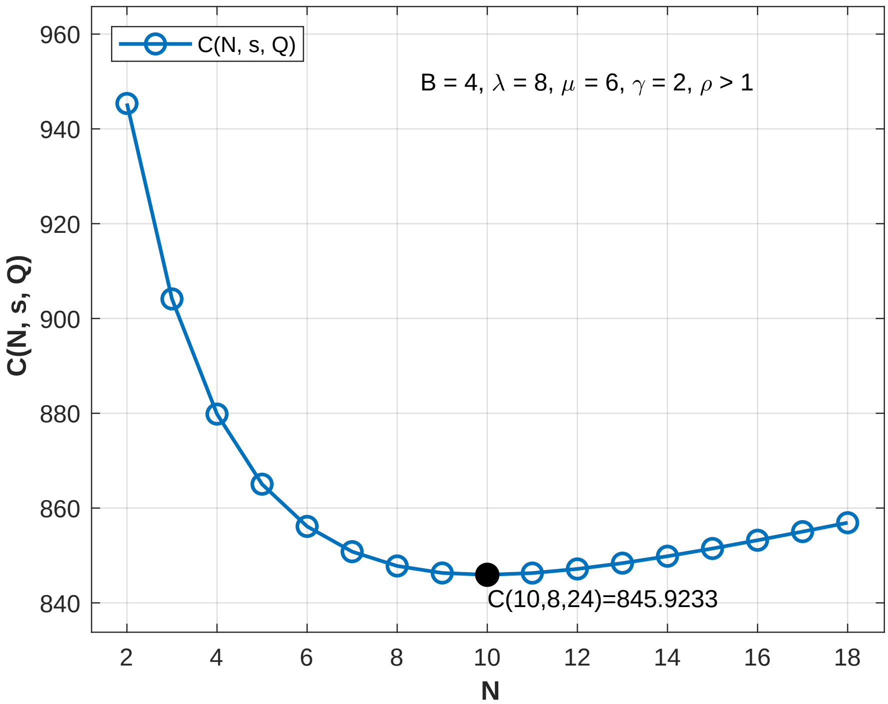


Figure 4. The convex representation of the cost value when $\rho > 1$

Table 9. The simulation-based comparison of the $M/M/1$ case with the $M/G/1$ type when $B = 4$

| | $\rho = \frac{\lambda}{\mu}$ | **Exponential distribution** $\mu = 6.0, \gamma = 2.0$ | **Deterministic distribution** $\mu = 6.0, \gamma = 2.0$ | **Normal distribution** $\mu = 6.0, \gamma = 2.0$ |
|---|---|---|---|---|
| $\lambda = 3.4$ | 0.57 | $C(30, 4, 11) = 214.44$ | $C(40, 3, 12) = 205.61$ | $C(47, 3, 13) = 208.33$ |
| $\lambda = 4.6$ | 0.77 | $C(48, 4, 18) = 270.88$ | $C(43, 4, 15) = 261.82$ | $C(45, 4, 16) = 264.90$ |
| $\lambda = 5.5$ | 0.92 | $C(27, 6, 22) = 379.43$ | $C(20, 6, 22) = 360.04$ | $C(23, 6, 22) = 368.23$ |
| $\lambda = 6.0$ | 1.0 | $C(20, 7, 24) = 462.05$ | $C(14, 6, 22) = 441.53$ | $C(16, 7, 23) = 449.21$ |
| $\lambda = 7.0$ | 1.17 | $C(12, 8, 23) = 651.36$ | $C(8, 7, 22) = 628.83$ | $C(9, 7, 22) = 638.18$ |
| $\lambda = 7.5$ | 1.25 | $C(12, 8, 23) = 748.09$ | $C(7, 8, 22) = 725.67$ | $C(9, 7, 23) = 730.09$ |
| $\lambda = 8.0$ | 1.33 | $C(9, 8, 22) = 844.02$ | $C(7, 8, 24) = 824.89$ | $C(8, 8, 23) = 826.77$ |
| $\lambda = 8.5$ | 1.42 | $C(9, 8, 24) = 942.39$ | $C(6, 7, 24) = 922.54$ | $C(7, 7, 25) = 925.12$ |
| $\lambda = 9.0$ | 1.5 | $C(7, 8, 25) = 1040.92$ | $C(7, 7, 25) = 1020.85$ | $C(7, 8, 25) = 1024.46$ |

Table 10. The simulation-based comparison of the $M/M/1$ case with the $M/G/1$ type when $B = 6$

| | $\rho = \frac{\lambda}{\mu}$ | **Exponential distribution** $\mu = 6.0, \gamma = 2.0$ | **Deterministic distribution** $\mu = 6.0, \gamma = 2.0$ | **Normal distribution** $\mu = 6.0, \gamma = 2.0$ |
|---|---|---|---|---|
| $\lambda = 3.4$ | 0.57 | $C(50, 6, 14) = 257.27$ | $C(45, 4, 10) = 246.72$ | $C(48, 5, 12) = 251.49$ |
| $\lambda = 4.6$ | 0.77 | $C(45, 6, 20) = 323.61$ | $C(40, 4, 17) = 313.88$ | $C(42, 5, 19) = 317.29$ |
| $\lambda = 5.5$ | 0.92 | $C(24, 7, 26) = 440.84$ | $C(17, 6, 22) = 420.44$ | $C(19, 6, 23) = 425.87$ |
| $\lambda = 6.0$ | 1.0 | $C(19, 7, 27) = 528.99$ | $C(13, 6, 23) = 506.66$ | $C(15, 6, 24) = 512.98$ |
| $\lambda = 7.0$ | 1.17 | $C(12, 8, 27) = 717.29$ | $C(10, 6, 24) = 692.70$ | $C(11, 6, 25) = 698.96$ |
| $\lambda = 7.5$ | 1.25 | $C(10, 8, 28) = 815.01$ | $C(9, 7, 25) = 791.81$ | $C(10, 7, 25) = 797.85$ |
| $\lambda = 8.0$ | 1.33 | $C(9, 8, 28) = 912.03$ | $C(8, 7, 25) = 890.05$ | $C(9, 7, 26) = 895.88$ |
| $\lambda = 8.5$ | 1.42 | $C(8, 8, 28) = 1010.77$ | $C(7, 7, 26) = 989.39$ | $C(7, 7, 27) = 994.88$ |
| $\lambda = 9.0$ | 1.5 | $C(7, 8, 28) = 1108.98$ | $C(7, 7, 27) = 1087.75$ | $C(7, 7, 27) = 1093.42$ |

Table 11. The simulation-based comparison of the $M/M/1$ case with the $M/G/1$ type when $B = 8$

| | $\rho = \frac{\lambda}{\mu}$ | **Exponential distribution** $\mu = 6.0, \gamma = 2.0$ | **Deterministic distribution** $\mu = 6.0, \gamma = 2.0$ | **Normal distribution** $\mu = 6.0, \gamma = 2.0$ |
|---|---|---|---|---|
| $\lambda = 3.4$ | 0.57 | $C(50, 8, 19) = 316.68$ | $C(44, 8, 16) = 305.84$ | $C(47, 8, 17) = 309.28$ |
| $\lambda = 4.6$ | 0.77 | $C(36, 8, 23) = 381.55$ | $C(31, 8, 20) = 369.26$ | $C(32, 8, 21) = 374.71$ |
| $\lambda = 5.5$ | 0.92 | $C(19, 8, 28) = 496.47$ | $C(13, 8, 25) = 477.81$ | $C(15, 8, 26) = 482.97$ |
| $\lambda = 6.0$ | 1.0 | $C(15, 8, 29) = 585.35$ | $C(10, 8, 26) = 562.96$ | $C(11, 8, 27) = 568.53$ |
| $\lambda = 7.0$ | 1.17 | $C(10, 8, 30) = 776.06$ | $C(9, 8, 27) = 754.22$ | $C(9, 8, 28) = 759.18$ |
| $\lambda = 7.5$ | 1.25 | $C(9, 8, 30) = 877.20$ | $C(8, 8, 28) = 855.41$ | $C(8, 8, 29) = 860.23$ |
| $\lambda = 8.0$ | 1.33 | $C(8, 8, 31) = 970.86$ | $C(7, 8, 28) = 948.98$ | $C(7, 8, 29) = 953.29$ |
| $\lambda = 8.5$ | 1.42 | $C(7, 8, 31) = 1067.33$ | $C(6, 8, 29) = 1045.38$ | $C(6, 8, 30) = 1050.71$ |
| $\lambda = 9.0$ | 1.5 | $C(6, 8, 32) = 1168.41$ | $C(5, 8, 30) = 1145.92$ | $C(6, 8, 30) = 1151.53$ |

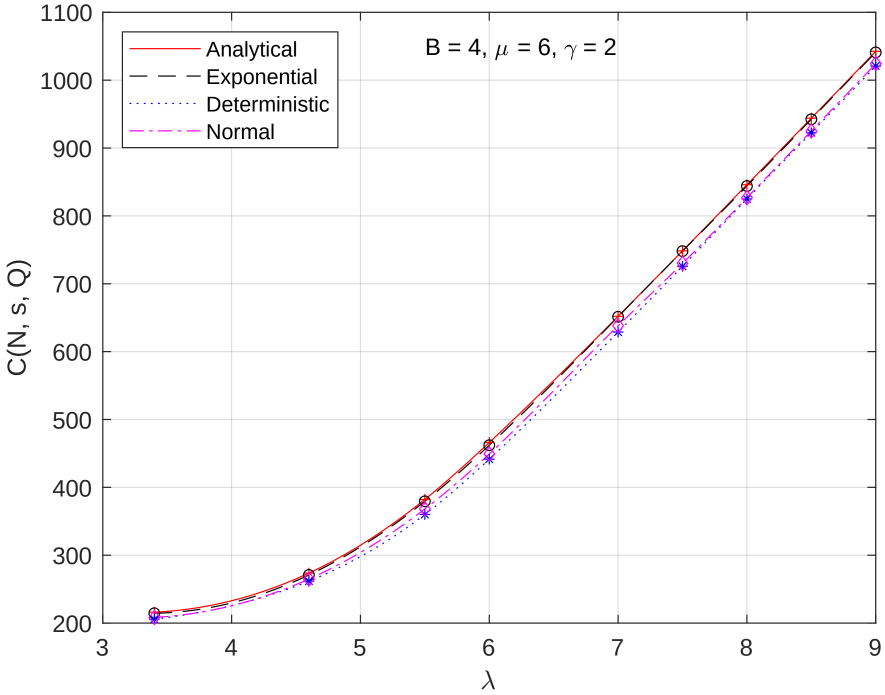


Figure 5. The cost behaviour of the analytical and simulation results for the $M/G/1$ type when $B = 4$

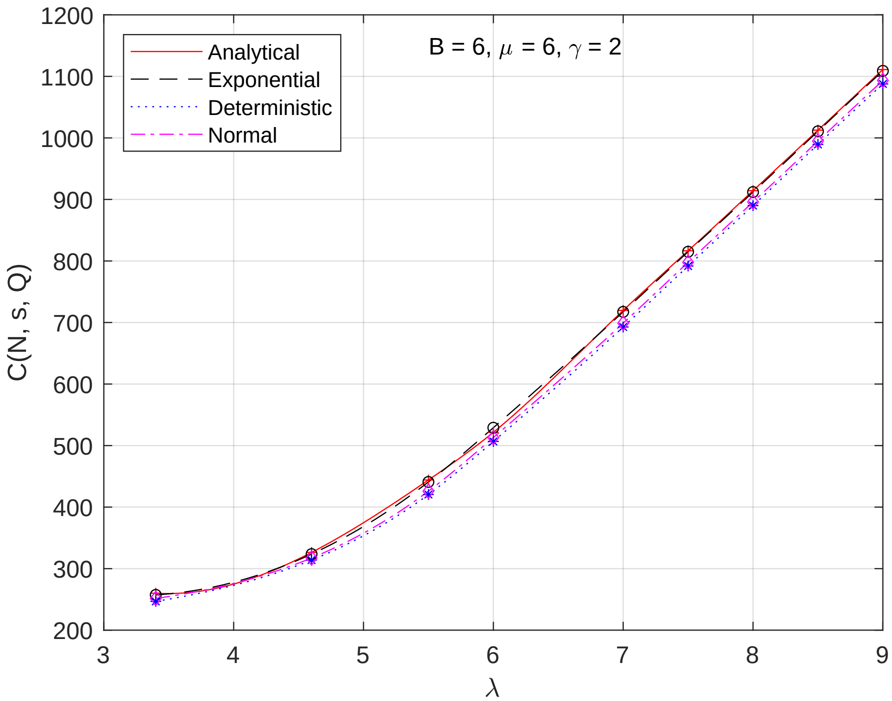


Figure 6. The cost behaviour of the analytical and simulation results for the $M/G/1$ type when $B = 6$

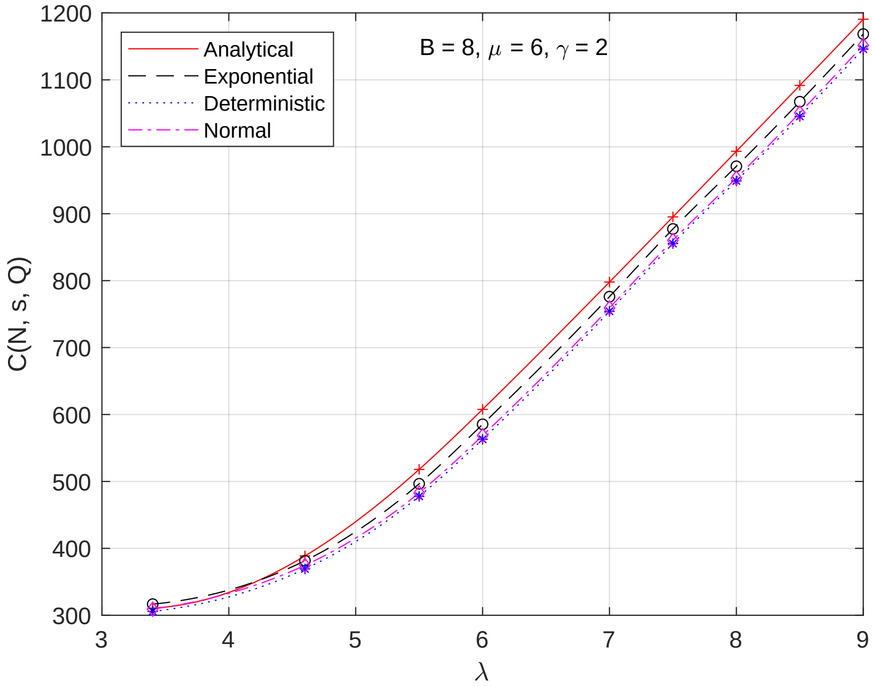


Figure 7. The cost behaviour of the analytical and simulation results for the $M/G/1$ type when $B = 8$

## 6. Conclusion

We investigated the $(s, Q)$ inventory system and batch demands in the $M/M/1$ queueing model with finite waiting space. We determined the steady-state joint probability distribution for customers in the system and the level of inventory in stock. The findings of this paper offer valuable implications by permitting customers to take multiple items in service. The explicit formula for relevant performance metrics are also derived. Numerical analysis is performed to emphasize the convex shape of the cost function. We made interesting observations of optimum waiting space, optimum order quantity and optimum order level, so that cost will be minimized. We verified our model by performing simulation results. Furthermore, we performed simulation results for the corresponding $M/G/1$ queueing model. After conducting extensive numerical studies, we gained valuable managerial insights and observed interesting outcomes in optimization of batch demand system. We believe that the study of the model described in this research may be helpful in a range of businesses as customers frequently demand for multiple items. For future studies, the model studied in this paper can be extended to relax the assumption of exponential service time and use a general service time distribution.

## Funding

The first and second authors acknowledge the National Board for Higher Mathematics (NBHM), DAE, Mumbai, India, for financial support from the project grant 02012/1/2019 NBHM(R.P)/R&D II/1231.

## Acknowledgments

The authors would like to express their gratitude to the Editor-in-Chief and the reviewer for their insightful comments, which helped to greatly improve the manuscript.


## References


[1] Berman, O., Kaplan, E. H., & Shevishak, D. G. (1993). Deterministic approximations for inventory management at service facilities. *IIE transactions*, 25(5), 98–104.

[2] Berman, O., & Kim, E. (1999). Stochastic models for inventory management at service facilities. *Stochastic Models*, 15(4), 695–718.

[3] Chakravarthy, S. R. (2020). Queueing-inventory models with batch demands and positive service times. *Automation and Remote Control*, 81(4), 713–730.

[4] Chakravarthy, S. R., & Hayat, K. (2020). Queueing-inventory models for a two-vendor system with positive service times. *Queueing Models and Service Management*, 3(1), 1–35.

[5] Chakravarthy, S. R., Maity, A., & Gupta, U. C. (2017). An $(s, S)$ inventory in a queueing system with batch service facility. *Annals of Operations Research*, 258, 263–283.

[6] Chakravarthy, S. R., & Rumyantsev, A. (2020). Analytical and simulation studies of queueing-inventory models with MAP demands in batches and positive phase type services. *Simulation Modelling Practice and Theory*, 103, 102092.

[7] Chakravarthy, S. R., & Subramanian, S. (2018). A stochastic model for automated teller machines subject to catastrophic failures and repairs. *Queueing Models and Service Management*, 1(1), 75–94.

[8] Divya, V., Krishnamoorthy, A., Vishnevsky, V. M., & Kozyrev, D. V. (2020). On a queueing system with processing of service items under vacation and $N$-policy with impatient customers. *Queueing Models and Service Management*, 3(2), 167–201.

[9] Karthick, T., Sivakumar, B., & Arivarignan, G. (2015). An inventory system with two types of customers and retrial demands. *International Journal of Systems Science: Operations & Logistics*, 2(2), 90-112.

[10] Karthikeyan, K., & Sudhesh, R. (2016). Recent review article on queueing inventory systems. *Research Journal of Pharmacy and Technology*, 9(11), 2056.

[11] Keerthana, M., Sivakumar, B., & Manuel, P. (2022). An inventory system with postponed and renewal demands. *International Journal of Systems Science: Operations & Logistics*, 9(2), 180–198.

[12] Krishnamoorthy, A., Lakshmy, B., & Manikandan, R. (2011). A survey on inventory models with positive service time. *Opsearch*, 48(2), 153–169.

[13] Krishnamoorthy, A., Shajin, D., & Narayanan, W. (2021). Inventory with positive service time: a survey. In: Anisimov, V., Limnios, N. (eds.), Queueing Theory 2, 201–237.

[14] Melikov, A., Krishnamoorthy, A., & Shahmaliyev, M. (2019). Numerical analysis and long run total cost optimization of perishable queuing inventory systems with delayed feedback. *Queueing Models and Service Management*, 2(1), 83-112.

[15] Saffari, M., Asmussen, S., & Haji, R. (2013). The $M/M/1$ queue with inventory, lost sale, and general lead times. *Queueing Systems*, 75(1), 65–77.

[16] Samanta, S. K., Isotupa, K. P. S., & Verma, A. (2023). Continuous review $(s, Q)$ inventory system at a service facility with positive order lead times. *Annals of Operations Research*, 331, 1007–1028.

[17] Schwarz, M., & Daduna, H. (2006). Queueing systems with inventory management with random lead times and with backordering. *Mathematical Methods of Operations Research*, 64(3), 383-414.

[18] Schwarz, M., Wichelhaus, C., & Daduna, H. (2007). Product form models for queueing networks with an inventory. *Stochastic Models*, 23(4), 627–663.

[19] Schwarz, M., Sauer, C., Daduna, H., Kulik, R., & Szekli, R. (2006). $M/M/1$ queueing systems with inventory. *Queueing Systems*, 54(1), 55–78.

[20] Sigman, K., & Simchi-Levi, D. (1992). Light traffic heuristic for an $M/G/1$ queue with limited inventory. *Annals of Operations Research*, 40(1), 371–380.

[21] Yue, D., Ye, Z., & Yue, W. (2023). Analysis of a queueing-inventory system with synchronous vacation of multiple servers. *Queueing Models and Service Management*, 6(1), 1-26.

[22] Yue, D., Zhao, G., & Qin, Y. (2018). An $M/M/1$ queueing-inventory system with geometric batch demands and lost sales. *Journal of Systems Science and Complexity*, 31(4), 1024–1041.