

Article

Multi-Server Queuing Production Inventory System with Emergency Replenishment

Dhanya Shajin ¹, Achyutha Krishnamoorthy ², Agassi Z. Melikov ³ and Janos Sztrik ^{4,*}¹ Department of Mathematics, S. N. College, Chempazhanchy 695587, Kerala, India² Centre for Research in Mathematics, CMS College, Kottayam 686001, Kerala, India³ Institute of Control Systems, National Academy of Science, AZ 1141 Baku, Azerbaijan⁴ Department of Informatics and Networks, Faculty of Informatics, University of Debrecen, 4032 Debrecen, Hungary

* Correspondence: sztrik.janos@inf.unideb.hu

Abstract: We consider a multi-server production inventory system with an unlimited waiting line. Arrivals occur according to a non-homogeneous Poisson process and exponentially distributed service time. At the service completion epoch, one unit of an item in the on-hand inventory decreases with probability δ , and the customer leaves the system without taking the item with probability $(1 - \delta)$. The production inventory system adopts an (s, S) policy where the processing of inventory requires a positive random amount of time. The production time for a unit item is phase-type distributed. Furthermore, assume that an emergency replenishment of one item with zero lead time takes place when the on-hand inventory level decreases to zero. The emergency replenishment is incorporated in the system to ensure customer satisfaction. We derive the stationary distribution of the system and some main performance measures, such as the distribution of the production on/off time in a cycle and the mean emergency replenishment cycle time. Numerical experiments are conducted to illustrate the system performance. A cost function is constructed, and we examine the optimal number of servers to be employed. Furthermore, we numerically calculate the optimal values of the production starting level and maximum inventory level.

Keywords: (s, S) production inventory system; non-homogeneous Poisson process; multi-server; emergency replenishment; cost function

MSC: 60K25; 90B22

Citation: Shajin, D.; Krishnamoorthy, A.; Melikov, A.Z.; Sztrik, J. Multi-Server Queuing Production Inventory System with Emergency Replenishment. *Mathematics* **2022**, *10*, 3839. <https://doi.org/10.3390/math10203839>

Academic Editors: Alexander Zeifman, Victor Korolev and Alexander Sipin

Received: 8 September 2022

Accepted: 12 October 2022

Published: 17 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Consider a c -server ($c \geq 2$) queuing inventory model with an inventory-level dependent Poisson arrival of customers. The service provided by each server has a duration that is exponentially distributed with parameter μ . On the completion of a service, the customer is served an item with probability δ . With probability $1 - \delta = \hat{\delta}$, the customer is not served the item. The replenishment of inventory is through a production process that is controlled by the (s, S) policy. To produce one unit of an item, the time required is phase-type distributed with representation (β, T) of order m . This production time can be interpreted as the time until the underlying Markov chain $\{\zeta(t), t \geq 0\}$, with a finite state space $\{1, 2, \dots, m, m + 1\}$, reaches the single absorbing state $m + 1$, conditioned on the fact that the initial state of this process is selected as one of the states $\{1, 2, \dots, m\}$ according to the initial probability vector $\beta = (\beta_1, \dots, \beta_m)$. The transition rates within the set $\{1, 2, \dots, m\}$ are defined by the generator T , and the transition rates into the absorbing state are given by $\mathbf{T}^0 = -T\mathbf{e}$.

The mean production of the item is calculated by $\eta' = -\beta T^{-1}\mathbf{e}$ (see Neuts [1]). All distributions involved are assumed to be independent of each other. When the on-hand inventory drops to zero, one unit of the same is purchased locally (emergency

replenishment) at a higher cost, to avoid customer loss. The time required for this is assumed to be negligible. This system is analyzed in the equilibrium state (a condition whose existence is established in the sequel). The objective of the proposed model is to minimize the total cost by determining the optimal number of servers and the optimal (s, S) policy.

As an example of the model: consider customers arriving at a car show room; a few of them decide to buy the car at the end of the advisors description of the vehicle. The remaining customers leave without buying the car. Further, as a justification for our assumption for an inventory-level dependent customer arrival rate, we notice the tendency among customers of products to queue up for an item when it is scarce with the hope that they will receive the item served within the next few replenishments. On the other end, we can cite the following situation that is commonly seen: when an item has a good brand name and is available without much of a delay, customers throng the shop where it is available, even when the waiting line is long.

Another example for the model described: If the item is abundantly available, customers are confident in receiving the item even when the queue size is large. On the other hand, if the item is scarce, naturally customers will be discouraged from joining the system. This is the reason for our assumption of an “inventory-level-dependent arrival rate” rather than “arrival rate” depending on the customers already present. In addition, it may be noted that if the arrival rate depends on the number of customers already present, then the present analysis needs drastic changes—either one has to resort to a truncation procedure to achieve a *LIQBD* process or go for analysis of the asymptotic quasi-Toeplitz Markov chain.

The rest of the article is organized as follows. Section 2 is the literature review. Section 3 provides our notations, assumptions and system analysis. Some important performance measures, including the distribution of the production on (or off) time and mean emergency replenishment time are also given in Section 3. Section 4 provides a numerical example, sensitivity analysis and some graphical illustrations. Finally, our conclusions are given in Section 5.

2. Literature Review

Queuing inventory has its origin in Melikov et al. [2] and Sigman and Levy [3]. These were followed by a spate of articles by Berman et al. [4–6]. In none of these did the authors attempt to derive the product-form solution. The first attempt to obtain a product-form solution was by Schwarz et al. [7] for the (s, S) , (r, Q) and random reorder point control policies.

The crucial assumption leading to the product-form solution is that, on inventory level dropping to zero during the replenishment lead time, no new customer is admitted to the system until the next replenishment materializes. Safari et al. [8] extends Schwarz et al. [7] to arbitrarily distributed lead time. Baek and Moon [9] is another paper providing a product-form solution. In their research paper, Krishnamoorthy and Raju (see [10]) introduced the concept of a local purchase in a (s, S) inventory system.

Yue et al. [11] investigated the results on a multi-server queuing inventory system with non-homogeneous Poisson arrivals, including the inventory-level-dependent arrival of customers. When the inventory level depletes to zero, one item is replenished immediately with zero lead time. They obtained a stability condition and computed the stationary distribution of the system. Furthermore, they derived the total cost function and illustrated numerical examples of an optimal (s, S) policy.

Haughton and Isotupa [12] analyzed an inventory system with lost sales with regular and emergency orders. They assumed that the lead time of regular orders and emergency orders were, respectively, stochastic and deterministic. There were higher costs per item for the emergency order. They compared the total costs for this system to a system without emergency order. Recently Yue and Qin [13] analyzed a production system with service time and a production vacation of random duration.

Refer the recent survey paper by Krishnamoorthy et al. [14] for more detailed reports on inventory with a positive service time and production-inventory system.

3. Model Description

3.1. Notations

In the sequel, we need the following notations and abbreviations:

- \mathbf{e} = Column vector of 1 s of appropriate order.
- $\mathbf{0}$ = Column vector of 0 s of appropriate order.
- I = Identity matrix of appropriate order.
- O = Zero matrix of appropriate order.
- *CTMC* : Continuous time Markov chain.
- *LIQBD* : Level independent quasi-birth and death process.
- *PH* : Phase-type distribution.
- *MAP* : Markovian Arrival Process.

3.2. Assumptions

Consider an (s, S) production inventory system with c -servers. Demands occur according to a non-homogeneous Poisson process with rate λ_j , which is dependent on the on-hand inventory level j , $1 \leq j \leq S$ and $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_S$. It takes an exponentially distributed time with parameter μ , to serve each customer. The on-hand inventory decreases by one unit with probability δ at the moment of a service completion and with probability $\hat{\delta} = (1 - \delta)$, the customer leaves the system without taking the item at the service completion epoch.

When the on-hand inventory depletes to s , the production process is switched *on*. Each production is of one unit and production process is kept in the on mode until the inventory level becomes S . It takes phase-type distributed with representation (β, T) of order m to produce one item. The mean production of the item is calculated by

$$\eta' = -\beta T^{-1} \mathbf{e}. \tag{1}$$

Further, each server serves independently of the others. The arrival, service and production processes are independent of each other.

In order to prevent the loss of demands when the inventory level equal to zero, an emergency replenishment policy is adopted in the system. When the on-hand inventory depletes to zero, one item is replenished immediately with zero lead time. Thus, there is no loss of demand in this model with the emergency replenishment policy in place. The production system is such that it takes negligible time for the item produced to reach the retail shop. Furthermore, we assume that $c < s$. The warehouse has limited capacity as we assume that S has an upper bound—say, M . Hence, $c < s < S \leq M$.

The structure of the system under study is given in Figure 1.

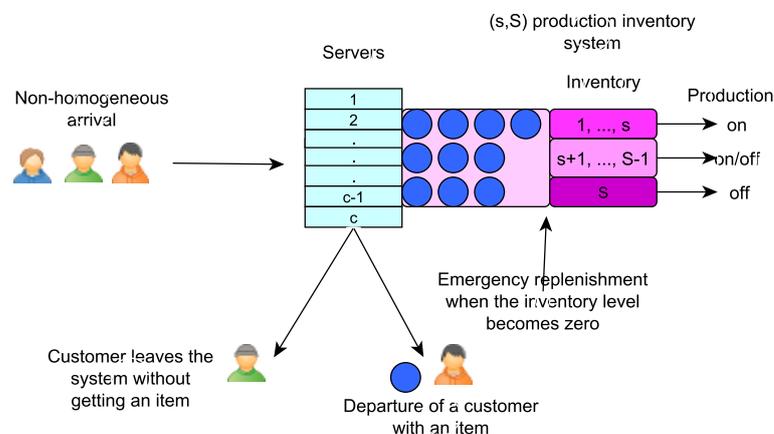


Figure 1. Picture representation of the model.

From the relation $\phi A = \mathbf{0}$, we have

$$\phi_i = \phi_1 \prod_{j=1}^{i-1} \mathcal{M}_j, 2 \leq i \leq S$$

where

$$\mathcal{M}_i = \begin{cases} -K[H_{i+1} + \mathcal{M}_{i+1}F_{i+2}]^{-1} & 1 \leq i \leq s-1 \\ -K_1[H_{i+1} + \mathcal{M}_{i+1}F_{i+2}]^{-1} & i = s \\ -K_2[H_{i+1} + \mathcal{M}_{i+1}F_{i+2}]^{-1} & s+1 \leq i \leq S-2 \\ -K_3H_{i+1}^{-1} & i = S-1 \end{cases}$$

with

$$F_i = (\delta\mu_{ci}I); 2 \leq i \leq s, \quad F_{s+1} = \begin{pmatrix} \delta\mu_{cs+1}\beta \\ \delta\mu_{cs+1}I \end{pmatrix},$$

$$F_i = \begin{pmatrix} \delta\mu_{ci} & \mathbf{0} \\ \mathbf{0} & \delta\mu_{ci}I \end{pmatrix}, \quad s+2 \leq i \leq S-1,$$

$$F_S = (\delta\mu_{cS} \quad \mathbf{0}),$$

$$K = (\mathbf{T}^0\beta), \quad K_1 = (0 \quad \mathbf{T}^0\beta), \quad K_2 = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^0\beta \end{pmatrix}, \quad K_3 = \begin{pmatrix} 0 \\ \mathbf{T}^0 \end{pmatrix},$$

$$H_1 = T, H_i = (T - \delta\mu_{ci}I), \quad 2 \leq i \leq s,$$

$$H_i = \begin{pmatrix} -\delta\mu_{ci} & \mathbf{0} \\ \mathbf{0} & (T - \delta\mu_{ci}I) \end{pmatrix}, \quad s+1 \leq i \leq S-1, H_S = \delta\mu_{cS}$$

From the relation (3), we have the stability condition as

$$\phi \left[\lambda_1 + \sum_{i=2}^S \lambda_i \prod_{j=1}^{i-1} \mathcal{M}_j \right] \mathbf{e} < \phi \left[\mu_{c1} + \sum_{i=2}^S \prod_{j=1}^{i-1} \mathcal{M}_j E_i + \sum_{i=2}^S \prod_{j=1}^{i-1} \mathcal{M}_j F_i \right] \mathbf{e} \tag{4}$$

with

$$E_i = \begin{pmatrix} \hat{\delta}\mu_{ci} & \mathbf{0} \\ \mathbf{0} & \hat{\delta}\mu_{ci}I \end{pmatrix}, s+1 \leq i \leq S-1, \quad E_i = \hat{\delta}\mu_{ci}I, 2 \leq i \leq s, \quad E_S = \hat{\delta}\mu_{cS}.$$

3.5. Steady State Probability Vector

Assuming that (3) is satisfied, we briefly outline the computation of the steady state system state probability. Let $\mathbf{y} = (\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \dots)$ be the steady-state probability vector of \mathcal{Q} . Then,

$$\mathbf{y}\mathcal{Q} = \mathbf{0}, \mathbf{y}\mathbf{e} = 1. \tag{5}$$

We see that \mathbf{y} , under the assumption that the stability condition (3) holds, is obtained as (see Theorem 3.1.1 of Neuts [1])

$$\mathbf{y}_{c-1+i} = \mathbf{y}_{c-1}R^i, \text{ for } i \geq 1 \tag{6}$$

where R is the minimal non-negative solution to the matrix quadratic equation

$$R^2A_2^{(c)} + RA_1^{(c)} + A_0 = O \tag{7}$$

and the boundary equations are given by

$$\begin{aligned} \mathbf{y}_0A_1^{(0)} + \mathbf{y}_1A_2^{(1)} &= \mathbf{0}, \\ \mathbf{y}_{i-1}A_0 + \mathbf{y}_iA_1^{(i)} + \mathbf{y}_{i+1}A_2^{(i+1)} &= \mathbf{0}, \quad 1 \leq i \leq c-2 \\ \mathbf{y}_{c-2}A_0 + \mathbf{y}_{c-1} \left[A_1^{(c-1)} + RA_2^{(c)} \right] &= \mathbf{0}. \end{aligned}$$

The normalizing condition $\mathbf{y}\mathbf{e} = 1$ gives

$$\mathbf{y}_0 \left[I + \sum_{i=1}^{c-2} \prod_{j=1}^i R_j + \prod_{j=1}^{c-1} R_j (I - R)^{-1} \right] \mathbf{e} = 1 \tag{8}$$

where

$$\begin{aligned} R_{c-1} &= -A_0 \left[A_1^{(c-1)} + RA_2^{(c)} \right]^{-1}, \\ R_i &= -A_0 \left[A_1^{(i)} + R_{i+1}A_2^{(i+1)} \right]^{-1}, \quad 1 \leq i \leq c-2. \end{aligned} \tag{9}$$

3.6. Performance Measures

In this section, we obtain expressions for a few additional system performance measures.

- The expected number of customers in the system:

$$E_N = \sum_{n=1}^{\infty} n \left[\sum_{i=1}^{S-1} \sum_{j=1}^m y_n(i, 1, j) + \sum_{i=s+1}^S y_n(i, 0) \right]. \tag{10}$$

- The expected number of items in the inventory:

$$E_I = \sum_{n=0}^{\infty} \left[\sum_{i=1}^{S-1} \sum_{j=1}^m i y_n(i, 1, j) + \sum_{i=s+1}^S i y_n(i, 0) \right]. \tag{11}$$

- The expected production rate:

$$E_{PR} = \frac{1}{\eta'} \sum_{n=0}^{\infty} \sum_{i=1}^{S-1} \sum_{j=1}^m y_n(i, 1, j) \tag{12}$$

where η' is given in (1).

- The expected rate at which the production process is switched on:

$$E_{on} = \delta\mu \left[\sum_{n=1}^c n y_n(s+1, 0) + c \sum_{n=c+1}^{\infty} y_n(s+1, 0) \right]. \tag{13}$$

- The expected rate at which the production process is switched off:

$$E_{off} = \frac{1}{\eta'} \sum_{n=0}^{\infty} \sum_{j=1}^m y_n(S-1, 1, j). \tag{14}$$

- The expected rate at which the emergency replenishment occurs:

$$E_{ER} = \delta\mu \sum_{n=1}^{\infty} \sum_{j=1}^m y_n(1, 1, j) \tag{15}$$

- The mean number of busy servers:

$$\begin{aligned} E_{BS} &= \sum_{n=1}^c \left[\sum_{i=1}^c \sum_{j=1}^m \min\{n, i\} y_n(i, 1, j) + n \left(\sum_{i=c+1}^{S-1} \sum_{j=1}^m y_n(i, 1, j) + \sum_{i=s+1}^S y_n(i, 0) \right) \right] \\ &+ \sum_{n=c+1}^{\infty} \left[\sum_{i=1}^c \sum_{j=1}^m \min\{c, i\} y_n(i, 1, j) + c \left(\sum_{i=c+1}^{S-1} \sum_{j=1}^m y_n(i, 1, j) + \sum_{i=s+1}^S y_n(i, 0) \right) \right]. \end{aligned} \tag{16}$$

with

$$\begin{aligned}
 a &= \left(\frac{c\delta\mu}{\eta}\right), \\
 \alpha_i &= \sum_{j=1}^{S-i} \left(\frac{c\delta\mu}{\eta}\right)^j, \quad s \leq i \leq S-1, \\
 \beta_i &= \prod_{j=i+1}^c \left(\frac{j\delta\mu}{\eta}\right), \quad 1 \leq i \leq c-1.
 \end{aligned}$$

Proof. Define $A = A_0 + A_1^{(c)} + A_2^{(c)}$.

Define $\pi = (\pi(1,1), \dots, \pi(s,1), \pi(s+1,0), \pi(s+1,1), \dots, \pi(S-1,0), \pi(S-1,1), \pi(S,0))$ as the steady state probability vector of the Markov chain whose infinitesimal generator is A . Then, we have

$$\pi A = \mathbf{0}, \quad \pi \mathbf{e} = 1.$$

From the relation $\pi A = \mathbf{0}$, we have

$$\begin{aligned}
 -\eta\pi(1,1) + 2\delta\mu\pi(2,1) &= 0, \\
 \eta\pi(i-1,1) - (\eta + i\delta\mu)\pi(i,1) + (i+1)\delta\mu\pi(i+1,1) &= 0, \quad 2 \leq i \leq c-1 \\
 \eta\pi(i-1,1) - (\eta + c\delta\mu)\pi(i,1) + c\delta\mu\pi(i+1,1) &= 0, \quad c \leq i \leq s-1 \\
 \eta\pi(s-1,1) - (\eta + c\delta\mu)\pi(s,1) + c\delta\mu(\pi(s+1,0) + \pi(s+1,1)) &= 0, \\
 -c\delta\mu\pi(i,0) + c\delta\mu\pi(i+1,0) &= 0, \quad s+1 \leq i \leq S-1 \\
 \eta\pi(i-1,1) - (\eta + c\delta\mu)\pi(i,1) + c\delta\mu\pi(i+1,1) &= 0, \quad s+1 \leq i \leq S-2 \\
 \eta\pi(S-2,1) - (\eta + c\delta\mu)\pi(S-1,1) &= 0, \\
 \eta\pi(S-1,1) - c\delta\mu\pi(S,0) &= 0.
 \end{aligned} \tag{20}$$

Solving the above set of equations, we obtain $\pi(i,1)$ for $1 \leq i \leq S-1$ and $\pi(i,0)$ for $s+1 \leq i \leq S-1$, which is expressed in (18).

The unknown probability $\pi(S,0)$ can be found from the normalizing condition

$$\sum_{i=1}^{S-1} \pi(i,1) + \sum_{i=s+1}^S \pi(i,0) = 1$$

and is expressed in (19).

The Markov chain is stable if and only if (see Theorem 3.1.1 of Neuts [1]) the left drift rate exceeds the right drift rate. Thus,

$$\pi A_0 \mathbf{e} < \pi A_2^{(c)} \mathbf{e}, \tag{21}$$

which implies the inequality (17). \square

3.7.2. Steady State Probability Vector

Assuming that (17) is satisfied, let $\mathbf{x} = (x_0, x_1, x_2, \dots)$ be the steady-state probability vector of \mathcal{Q} . Then,

$$\mathbf{x}\mathcal{Q} = \mathbf{0}, \quad \mathbf{x}\mathbf{e} = 1. \tag{22}$$

Under the assumption that the stability condition (17) holds,

$$\mathbf{x}_{c-1+i} = \mathbf{x}_{c-1} R^i, \quad \text{for } i \geq 1 \tag{23}$$

where R is the minimal non-negative solution to the matrix quadratic equation

$$R^2 A_2^{(c)} + R A_1^{(c)} + A_0 = O \tag{24}$$

and the boundary equations are given by

$$\begin{aligned} \mathbf{x}_0 A_1^{(0)} + \mathbf{x}_1 A_2^{(1)} &= \mathbf{0}, \\ \mathbf{x}_{i-1} A_0 + \mathbf{x}_i A_1^{(i)} + \mathbf{x}_{i+1} A_2^{(i+1)} &= \mathbf{0}, \quad 1 \leq i \leq c-2 \\ \mathbf{x}_{c-2} A_0 + \mathbf{x}_{c-1} [A_1^{(c-1)} + R A_2^{(c)}] &= \mathbf{0}. \end{aligned}$$

The normalizing condition $\mathbf{x}\mathbf{e} = 1$ gives

$$\mathbf{x}_0 \left[I + \sum_{i=1}^{c-2} \prod_{j=1}^i R_j + \prod_{j=1}^{c-1} R_j (I - R)^{-1} \right] \mathbf{e} = 1 \tag{25}$$

where

$$\begin{aligned} R_{c-1} &= -A_0 [A_1^{(c-1)} + R A_2^{(c)}]^{-1}, \\ R_i &= -A_0 [A_1^{(i)} + R_{i+1} A_2^{(i+1)}]^{-1}, \quad 1 \leq i \leq c-2. \end{aligned} \tag{26}$$

Next, we proceed to derive certain system characteristics that shed light on its performance. Given $\epsilon > 0$, there exists an \mathbf{N} sufficiently large such that

$$\sum_{n=0}^{\mathbf{N}} \mathbf{x}_n \mathbf{e} > 1 - \epsilon.$$

Except for heavy traffic, the above approximation is near to the exact value.

3.7.3. Analysis of the Emergency Replenishment Cycle

The emergency cycle time T is the time length between two adjacent emergency replenishment epochs. Once the last remaining item in the inventory is taken away, an emergency replenishment will occur. Thus, a new item will be instantly purchased and brought to the inventory because it is assumed that there is no lead time for an emergency replenishment. Therefore, the state $I(t) = 0$ can be seen as a virtual (instantaneous) state.

Now, consider the Markov chain $\{(N(t), I(t), C(t)), t \geq 0\}$ with state space $\{(n, i, 1), 0 \leq n \leq \mathbf{N}, 1 \leq i \leq S-1\} \cup \{(n, i, 0), 0 \leq n \leq \mathbf{N}, s+1 \leq i \leq S\} \cup \{0\}$ where $\{0\}$ denotes the absorbing state. Thus, an emergency replenishment cycle time T is the first passage time from the state $(n, 1, 1), 0 \leq n \leq \mathbf{N}$ to the absorbing state $\{0\}$.

Let $\alpha_N = (\xi_0, \xi_1, \dots, \xi_N)$ where

$$\xi_n = \frac{(x_n(1, 1), 0, \dots, 0)}{\sum_{n=0}^{\mathbf{N}} x_n(1, 1)}, 0 \leq n \leq \mathbf{N}.$$

This can be taken as the initial probability vector of an emergency replenishment cycle time T . Thus, there is only one item at the beginning of an emergency replenishment cycle time T . The first passage time from the state $(n, 1, 1)$ to $\{0\}$ follows a phase-type distribution with representation $(\alpha_N, \mathcal{W}_N)$. The corresponding infinitesimal generator is of the form $\begin{pmatrix} \mathcal{W}_N & \mathbf{W}_N^0 \\ \mathbf{0} & 0 \end{pmatrix}$ where \mathcal{W}_N is a matrix of order $(\mathbf{N} + 1)(2S - (s + 1))$ and is given by

$$E_2^{(n)} = \begin{pmatrix} \delta n\mu & & & \\ \delta n\mu & \delta n\mu & & \\ & \ddots & \ddots & \\ & & \delta n\mu & \delta n\mu \end{pmatrix}, 1 \leq n \leq c$$

$$E_1^{(0)} = \begin{pmatrix} -\lambda_{s+1} & & & \\ & -\lambda_{s+2} & & \\ & & \ddots & \\ & & & -\lambda_S \end{pmatrix}, \tilde{E}_1^{(c)} = \begin{pmatrix} -c\mu & & & \\ & -c\mu & & \\ & & \ddots & \\ & & & -c\mu \end{pmatrix}.$$

The production *off* cycle time T follows a phase-type distribution with representation (γ_N, Z_N) where $\gamma_N = (\tau_0, \tau_1, \dots, \tau_N)$ with

$$\tau_n = \frac{(0, \dots, 0, x_n(S-1, 1))}{\sum_{n=0}^N x_n(S-1, 1)}, \text{ for } 0 \leq n \leq N.$$

Thus, the expected length of production *off* cycle T is given by

$$E_T^{(N)}(off) = -\gamma_N Z_N^{-1} \mathbf{e}.$$

3.7.6. Additional Performance Measures

In this section, we obtain expressions for a few additional system performance measures.

- The expected number of customers in the system:

$$E_N = \sum_{n=1}^{\infty} n \left[\sum_{i=1}^{S-1} x_n(i, 1) + \sum_{i=s+1}^S x_n(i, 0) \right]. \tag{30}$$

- The expected number of items in the inventory:

$$E_I = \sum_{n=0}^{\infty} \left[\sum_{i=1}^{S-1} i x_n(i, 1) + \sum_{i=s+1}^S i x_n(i, 0) \right]. \tag{31}$$

- The expected production rate:

$$E_{PR} = \eta \sum_{n=0}^{\infty} \sum_{i=1}^{S-1} x_n(i, 1). \tag{32}$$

- The expected rate at which the production process is switched on:

$$E_{on} = \mu\delta \left[\sum_{n=1}^c n x_n(s+1, 0) + c \sum_{n=c+1}^{\infty} x_n(s+1, 0) \right]. \tag{33}$$

- The expected rate at which the production process is switched off:

$$E_{off} = \eta \sum_{n=0}^{\infty} x_n(S-1, 1). \tag{34}$$

- The average emergency replenishment rate:

$$E_{ER} = \frac{1}{E_T} \sum_{n=1}^{\infty} x_n(1, 1) \tag{35}$$

where E_T is defined in Section 3.7.3.

- The mean number of busy servers:

$$E_{BS} = \sum_{n=1}^c \left[\sum_{i=1}^c \min\{n, i\} x_n(i, 1) + n \left(\sum_{i=c+1}^{S-1} x_n(i, 1) + \sum_{i=s+1}^S x_n(i, 0) \right) \right] + \sum_{n=c+1}^{\infty} \left[\sum_{i=1}^c \min\{c, i\} x_n(i, 1) + c \left(\sum_{i=c+1}^{S-1} x_n(i, 1) + \sum_{i=s+1}^S x_n(i, 0) \right) \right]. \tag{36}$$

3.8. An Optimization Problem

Based on the above performance measures, we constructed a cost function for determining the optimality of s, S and c .

Consider the following cost parameters:

- h_{cust} : holding cost of a customer per unit time.
- h_{inv} : holding cost of an item per unit time.
- C_{PR} : cost of ‘production on’ per unit time.
- K : fixed cost for starting the production.
- C_{ER} : cost of each emergency replenishment.
- C_{idle} : cost incurred per idle server per time.
- C_{busy} : cost incurred per busy server per time.

Now, we have the cost per unit time given by

$$\mathcal{F}(c, s, S) = h_{cust}E_N + h_{inv}E_I + C_{PR}E_{PR} + KE_{on} + C_{ER}E_{ER} + C_{idle}(c - E_{BS}) + C_{busy}E_{BS}$$

where $E_N, E_I, E_{PR}, E_{on}, E_{ER}$ and E_{BS} are given in Sections 3.6 and 3.7.6.

4. Numerical Illustrations

Next, we proceed to a few numerical examples in order to bring out the system behavior with respect to certain parameters. Here, demands occur according to a non-homogeneous Poisson process with rate λ_j , which depends on the on-hand inventory level $j; 1 \leq j \leq S$ and $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_S$. We use

$$\lambda_j = \lambda j^\gamma, 1 \leq j \leq S \tag{37}$$

where $0 \leq \gamma \leq 1$ (see Alfares [15]). This is an increasing function of the on-hand inventory level. If $\gamma = 0$, then the demand rate is homogeneous Poisson process with rate λ .

Choose $\gamma = 0.1$ in the following examples.

4.1. Example: Effect of λ

From Table 1, for the indicated input parameters, we note that the expected number of customers in the system increases with the increase in the arrival rate. This results in the fast depletion of inventory since the service rate is high. In turn, this results in an increased production rate since the production unit has to be on for a longer time duration. As a consequence, the emergency replenishment rate E_{ER} increases, and the mean number of busy servers also increases.

Table 1. Effect of λ for $(S, s, c, \delta, \mu, \eta) = (35, 12, 5, 0.8, 7, 2.6)$.

λ	E_N	E_I	E_{PR}	E_{on}	E_{ER}	E_{BS}
1	0.1949	23.2314	1.0913	0.0315	0.1949	0.0001
1.5	0.2912	22.4978	1.6307	0.0291	0.2912	0.0002
2	0.3840	20.5180	2.1476	0.0173	0.3838	0.0003
2.5	0.4611	14.2006	2.5228	0.0035	0.4576	0.0071
3	0.5341	6.9958	2.5974	0.0007	0.5065	0.0425
3.5	0.6433	4.3116	2.5981	0.0002	0.5630	0.0975

4.2. Example: Effect of μ

A faster service rate ensures a reduced queue size and also a reduced inventory level (see Table 2). This, in turn, makes the production unit work over a longer duration in a cycle resulting in an increased production rate. Thus, emergency ordering rate drastically decreases as well as the mean number of busy servers.

Table 2. Effect of μ for $(S, s, c, \delta, \lambda, \eta) = (35, 15, 5, 0.8, 6, 2.6)$.

μ	E_N	E_I	E_{PR}	E_{on}	E_{ER}	E_{BS}
5	2.5101	6.0240	2.4648	0.0171	1.2564	0.3058
6	2.0056	4.7143	2.5076	0.0133	1.0498	0.3308
7	1.6307	3.6792	2.5411	0.0093	0.8999	0.3365
8	1.3458	2.9702	2.5641	0.0061	0.7871	0.3271
9	1.1296	2.5288	2.5785	0.0038	0.6995	0.3094
10	0.9651	2.2678	2.5871	0.0024	0.6295	0.2888

4.3. Example: Effect of η

Table 3: An increased production rate ensures an increase in the availability of inventory and hence a faster disposal of demands. It also results in an increased rate of production and reduced ‘switching on’ of the production mechanism. However, reduced ‘switching on’ results in an increase in emergency replenishment. This, in turn, ensures a reduction in the number of busy servers.

Table 3. Effect of η for $(S, s, c, \delta, \lambda, \mu) = (33, 15, 3, 0.8, 5, 7)$.

η	E_N	E_I	E_{PR}	E_{on}	E_{ER}	E_{BS}
1.4	1.2910	4.1961	1.3475	0.0154	0.7459	0.3420
1.6	1.2559	4.2002	1.5435	0.0145	0.7498	0.3216
1.8	1.2223	4.2163	1.7404	0.0136	0.7539	0.3011
2	1.1899	4.2468	1.9380	0.0127	0.7582	0.2805
2.2	1.1590	4.2949	2.1364	0.0118	0.7628	0.2598
2.4	1.1294	4.3647	2.3354	0.0109	0.7678	0.2390

4.4. Example: Effect of δ

Table 4: δ is the probability of a customer being served one unit of inventoried item at the end of their service. Interestingly, an increase in the value of δ initially results in a decrease in the expected number of customers and then starts increasing with an increase in the value of δ . Nevertheless, the expected inventory level continues decreasing drastically with the increase in the value of δ . This is no surprise since the probability of a customer served the inventoried item results in a decrease in the expected items held. The expected production rate (E_{PR}) has to increase. However, E_{on} decreases first and then increases because the production process is over a longer duration. A reasonably high production rate ensures reduced emergency purchases. The mean number of busy servers continues increasing.

Table 4. Effect of δ for $(S, s, c, \eta, \lambda, \mu) = (40, 15, 3, 2.6, 4, 7)$.

δ	E_N	E_I	E_{PR}	E_{on}	E_{ER}	E_{BS}
0.1	0.8100	27.7074	0.5731	0.0267	0.7918	0.0001
0.2	0.8082	27.1839	1.1214	0.0250	0.7902	0.0001
0.3	0.8050	26.3072	1.6633	0.0241	0.7873	0.0003
0.4	0.7957	23.9226	2.1781	0.0148	0.7783	0.0004
0.5	0.7620	15.9770	2.5265	0.0037	0.7393	0.0087
0.6	0.7388	8.0777	2.5752	0.0020	0.6814	0.0481
0.7	0.7534	5.4415	2.5717	0.0028	0.6504	0.0969
0.8	0.7776	4.3754	2.5681	0.0037	0.6331	0.1403
0.9	0.8020	3.8054	2.5659	0.0044	0.6219	0.1770
1	0.8247	3.4446	2.5647	0.0051	0.6139	0.2079

4.5. Example: Effect of s and S on Various Performance Measures

Table 5 indicates that, by increasing the maximum level of inventory, the expected number of customers can be brought down. This is because an increase in the inventory level results in a larger number of services. On the other hand, an increase in the value of S results in an increase in the number of items held in the inventory (see Table 5, Figures 2 and 3).

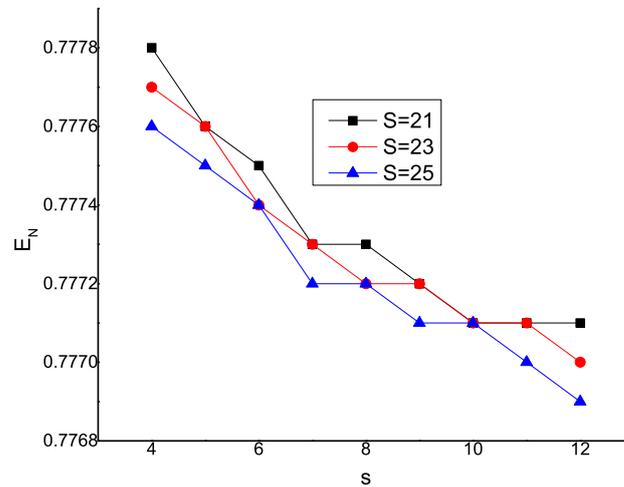


Figure 2. Effect on E_N .

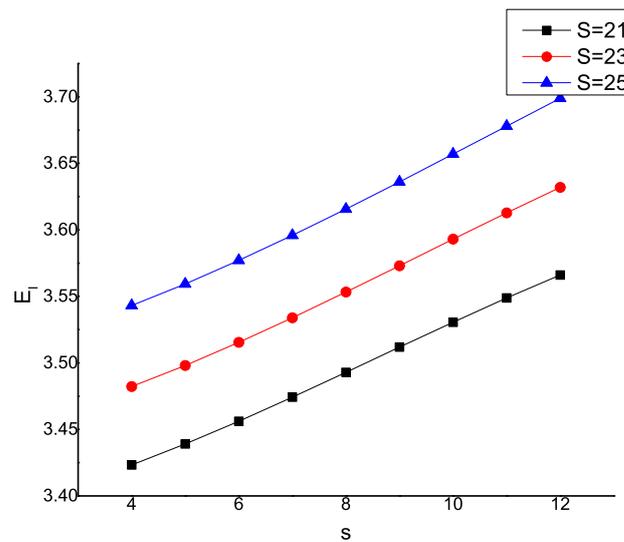


Figure 3. Effect on E_I .

Table 5. Effect of s and S for $(c, \delta, \lambda, \mu, \eta) = (3, 0.8, 4, 7, 2.6)$.

Effect on E_N				Effect on E_I			
s	$S = 21$	$S = 23$	$S = 25$	s	$S = 21$	$S = 23$	$S = 25$
4	0.7778	0.7777	0.7776	4	3.4233	3.4822	3.5432
5	0.7776	0.7776	0.7775	5	3.4391	3.4981	3.5593
6	0.7775	0.7774	0.7774	6	3.4561	3.5154	3.577
7	0.7773	0.7773	0.7772	7	3.4742	3.5339	3.5958
8	0.7773	0.7772	0.7772	8	3.4929	3.5532	3.6156
9	0.7772	0.7772	0.7771	9	3.5118	3.573	3.6361
10	0.7771	0.7771	0.7771	10	3.5306	3.593	3.657
11	0.7771	0.7771	0.7770	11	3.5489	3.6128	3.678
12	0.7771	0.7770	0.7769	12	3.566	3.632	3.6989

Table 6 are indications of the fact that the production rate increases with an increase in the value of s . This is the consequences of the early start of production. However, the same thing brings down the duration of the ‘production on’ status (see Table 6). For a fixed value of S and varying s , E_{PR} (see Figure 4) increases and E_{on} (see Figure 5) decreases. But for a fixed value of s and varying S , both E_{PR} (see Figure 4) and E_{on} (see Figure 5) decrease. These are consequences of the delayed ‘switching on’ of production.

Table 6. Effect of s and S for $(c, \delta, \lambda, \mu, \eta) = (3, 0.8, 4, 7, 2.6)$.

Effect on E_{PR}				Effect on E_{on}			
s	$S = 21$	$S = 23$	$S = 25$	s	$S = 21$	$S = 23$	$S = 25$
4	2.5681	2.5659	2.5634	4	0.0042	0.0041	0.004
5	2.5707	2.5684	2.566	5	0.0041	0.004	0.0039
6	2.5732	2.5709	2.5684	6	0.004	0.0039	0.0038
7	2.5756	2.5733	2.5708	7	0.0039	0.0038	0.0037
8	2.5779	2.5756	2.573	8	0.0038	0.0037	0.0036
9	2.5801	2.5778	2.5752	9	0.0037	0.0036	0.0035
10	2.5822	2.5799	2.5774	10	0.0036	0.0035	0.0034
11	2.5843	2.5819	2.5794	11	0.0035	0.0034	0.0033
12	2.5862	2.5839	2.5814	12	0.0034	0.0033	0.0032

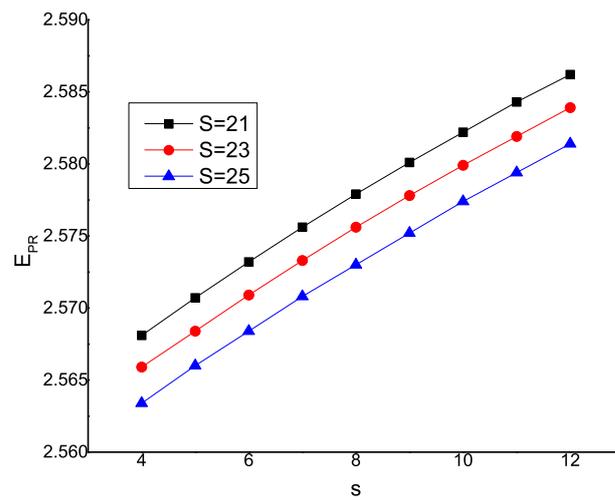


Figure 4. Effect on E_{PR} .

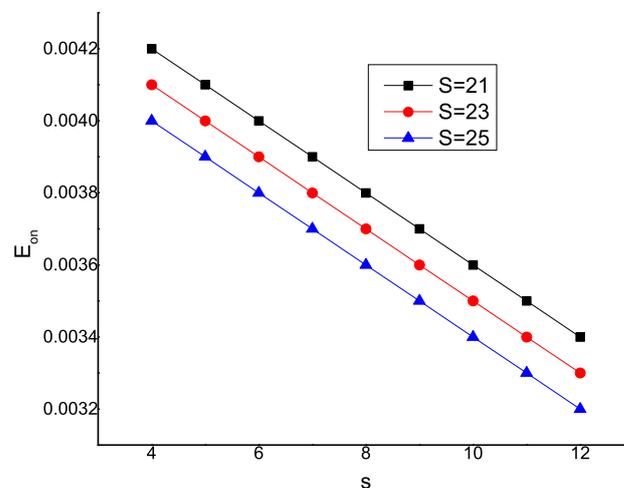


Figure 5. Effect on E_{on} .

Table 7 provides the expected number of busy servers E_{BS} (see Figure 6) with respect to s and S and the expected emergency replenishment values E_{ER} (see Figure 7). For fixed s and varying (increasing) S , E_{BS} and E_{ER} , respectively, increase/decrease, which is a consequence of the greater availability of inventoried items.

Table 7. Effect of s and S for $(c, \delta, \lambda, \mu, \eta) = (3, 0.8, 4, 7, 2.6)$.

Effect on E_{BS}				Effect on E_{ER}			
s	$S = 21$	$S = 23$	$S = 25$	s	$S = 21$	$S = 23$	$S = 25$
4	0.6249	0.6255	0.6261	4	0.1494	0.1488	0.1481
5	0.6252	0.6257	0.6263	5	0.149	0.1483	0.1477
6	0.6254	0.626	0.6265	6	0.1486	0.1479	0.1473
7	0.6256	0.6262	0.6268	7	0.1482	0.1475	0.1469
8	0.6259	0.6265	0.627	8	0.1479	0.1472	0.1465
9	0.6261	0.6267	0.6273	9	0.1475	0.1469	0.1462
10	0.6263	0.6269	0.6275	10	0.1473	0.1466	0.1459
11	0.6265	0.6271	0.6277	11	0.1470	0.1463	0.1456
12	0.6267	0.6273	0.6279	12	0.1468	0.1461	0.1454

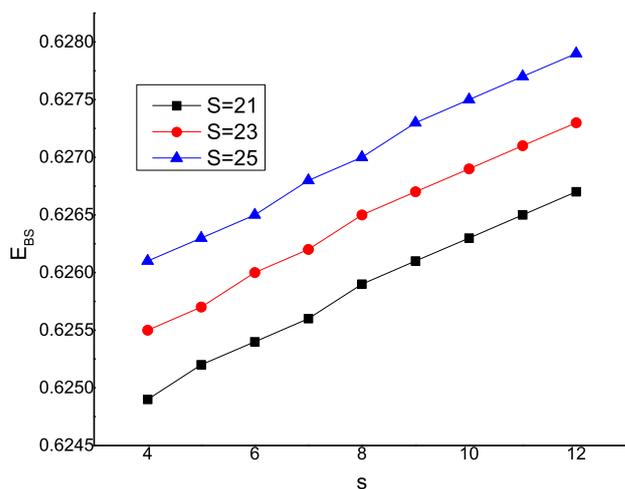


Figure 6. Effect on E_{BS} .

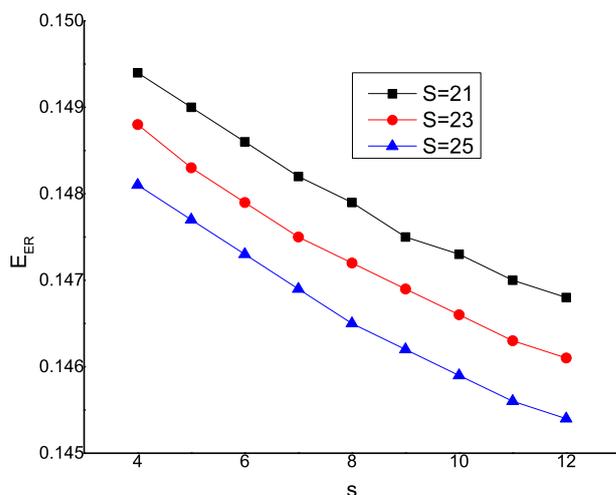


Figure 7. Effect on E_{ER} .

4.6. Example: Optimal Cost

In order to study the variation in different parameters on the cost function described in Section 3.8, we first take the values $(K, h_{cust}, h_{inv}, C_{PR}, C_{ER}, C_{idle}, C_{busy}) = (\$2000, \$0.5, \$5, \$100, \$250, \$40, \$5)$.

Now, we numerically compute the optimal (s, S) pair (denoted by (s^*, S^*)) such that

$$\mathcal{F}(c, s^*, S^*) = \min_{s, S} \mathcal{F}(c, s, S). \tag{38}$$

4.6.1. Example (a)

From Table 8, the optimal pair $(s^*, S^*) = (6, 16)$ and optimum cost $\mathcal{F}(c, s^*, S^*) = \606.0135 for the input parameters $\gamma = 0.1, c = 3, \lambda = 4, \mu = 7, \delta = 0.8$ and $T = \begin{pmatrix} -4 & 0 & 1 \\ 3 & -3 & 1 \\ 2 & 1 & -5 \end{pmatrix}, \beta = (1 \ 0 \ 0)$.

Table 8 provides the optimal running cost of the system for the indicated input parameters. The reason for the increase in cost for pairs (s, S) for fixed S while s is varied is evident: the increased holding cost. However, with S increasing, we notice that the system cost decreases drastically. This is attributed to reduced emergency purchases, reduced production cost per unit time and, to some extent, reduced holding costs.

Table 8. $\mathcal{F}(c, s, S)$.

s	$S = 6$	$S = 7$	$S = 8$	$S = 9$	$S = 10$	$S = 11$	$S = 12$	$S = 13$	$S = 14$	$S = 15$	$S = 16$
4	910.4632	769.3924	702.8011	665.8797	643.6297	629.5764	620.4579	614.4529	610.47	607.8226	606.065
5		812.0338	715.8488	670.4187	645.2558	630.1284	620.6088	614.4614	610.4358	607.7826	606.0311
6			745.3399	679.4886	648.4402	631.2916	621.022	614.5905	610.4605	607.773	606.0135
7				699.7053	654.6623	633.4851	621.8327	614.887	610.5606	607.7993	606.0141
8					668.4096	637.7087	623.3235	615.4421	610.7679	607.8731	606.0369
9						646.9861	626.1649	616.4447	611.1432	608.0154	606.0896
10							632.3806	618.341	611.8114	608.2665	606.186
11								622.4769	613.0679	608.7086	606.3526
12									615.8021	609.5356	606.643
13										611.332	607.1841
14											608.3575

4.6.2. Example (b)

From Table 9, the optimal pair $(s^*, S^*) = (10, 16)$ and optimum cost $\mathcal{F}(c, s^*, S^*) = \419.8059 for the indicated input parameters.

Table 9. $\mathcal{F}(c, s, S)$ for $(c, \eta, \lambda, \mu, M) = (3, 2.6, 4, 7, 16)$.

s	$S = 6$	$S = 7$	$S = 8$	$S = 9$	$S = 10$	$S = 11$	$S = 12$	$S = 13$	$S = 14$	$S = 15$	$S = 16$
4	644.1595	533.2389	483.2442	456.9492	441.992	433.105	427.6854	424.3232	422.2076	420.8558	419.9742
5		566.9304	493.9666	461.0212	443.692	433.846	428.0056	424.4485	422.2407	420.8466	419.9479
6			516.4134	468.1623	446.3843	434.9474	428.464	424.6268	422.2922	420.8412	419.9197
7				483.0568	451.1004	436.6985	429.1542	424.8899	422.3721	420.8416	419.888
8					460.9622	439.7929	430.2733	425.3026	422.5029	420.8549	419.8542
9						446.3231	432.2923	426.0012	422.7302	420.8969	419.8229
10							436.6286	427.3109	423.1505	421.001	419.8059
11								430.2076	423.9936	421.2372	419.8279
12									425.9467	421.7731	419.9422
13										423.1067	420.2747
14											421.1995

Table 9 provides the optimal running cost of the system for the indicated input parameters. The reason for the increase in cost for pairs (s, S) for fixed S while s is varied is

evident: the increased holding cost. However, with S increasing, we notice that the system cost decreases drastically. This is attributed to reduced emergency purchases, reduced production cost per unit time and, to some extent, reduced holding costs.

Now, we compute the optimal (s^*, S^*) and optimal server c^* such that

$$\mathcal{F}(c^*, s^*, S^*) = \min_{c, s, S} \mathcal{F}(c, s, S). \tag{39}$$

At the optimal pair $(s^*, S^*) = (10, 16)$, the minimum cost $\mathcal{F}(c^*, s^*, S^*) = \390.7507 is obtained numerically when $c^* = 2$.

4.7. Example: The relation between δ and Cost

In order to study the variation in different parameters on cost function, we first take the values $(K, h_{cust}, h_{inv}, C_{PR}, C_{ER}, C_{idle}, C_{busy}) = (\$5000, \$0.05, \$5, \$100, \$200, \$30, \$5)$.

Table 10 provides the effect of the probability of an item being served to a customer on the cost function for the input parameters $S = 13, s = 6, \gamma = 0.1, \lambda = 5, \mu = 6, T =$

$$\begin{pmatrix} -4 & 1 & 3 \\ 3 & -3 & 1 \\ 2 & 1 & -8 \end{pmatrix}, \beta = (0.2 \ 0.5 \ 0.3).$$

From Figure 8, we found that the increase in the value of δ results initially decreases $\mathcal{F}(c, s, S)$ and then starts increasing with the increase in the value of δ . We obtained the minimum cost when $\delta = 0.2$.

Table 10. Effect of δ on $\mathcal{F}(c, s, S)$.

δ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\mathcal{F}(2, s, S)$	246.0757	195.9772	280.9794	376.3558	473.1766	570.4884	668.0241	765.6783	863.4014	961.1673
$\mathcal{F}(3, s, S)$	274.6499	224.8902	310.326	406.0481	503.1662	600.737	698.5022	796.3634	894.2765	992.2195
$\mathcal{F}(4, s, S)$	304.1992	254.5151	340.1136	436.0469	533.3815	631.1632	729.1311	827.1869	925.2874	1023.4

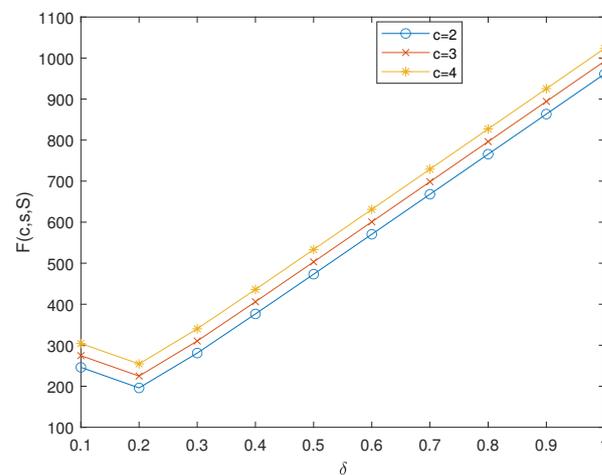


Figure 8. Effect of δ on $\mathcal{F}(c, s, S)$.

4.8. Example: The Optimal Number of Servers and Minimum Cost

In order to study the variation in different parameters on the cost function, we first take the values $(K, h_{cust}, h_{inv}, C_{PR}, C_{ER}, C_{idle}, C_{busy}) = (\$2000, \$0.5, \$5, \$100, \$125, \$3, \$2)$.

Table 11 provides the optimal value of the servers to be employed for the given input parameters. We look at the effects of the probability of an item being served to a customer on the number of servers to be employed. When the probability of an item being served to the customer is equal to at least half, then the optimal number of servers to be employed is stabilized, and the system cost is also drastically cut.

Table 11. Optimal server c^* and corresponding minimum cost for $(S, s, \eta, \lambda, \mu) = (30, 12, 2.6, 5, 7)$.

δ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
c^*	2	3	4	4	5	5	5	5	5	5
$\mathcal{F}(c^*, s, S)$	445.5808	477.6176	419.2914	349.7553	318.587	318.0415	322.3932	326.9127	330.8801	334.2469

4.9. Example: The Relation between δ and E_N

In this example, we give a comparison between the constant arrival case (λ) and inventory-level-dependent arrival case ($\lambda_i, 1 \leq i \leq S$) in Figure 9 and found the two cases to be showing entirely different behaviors.

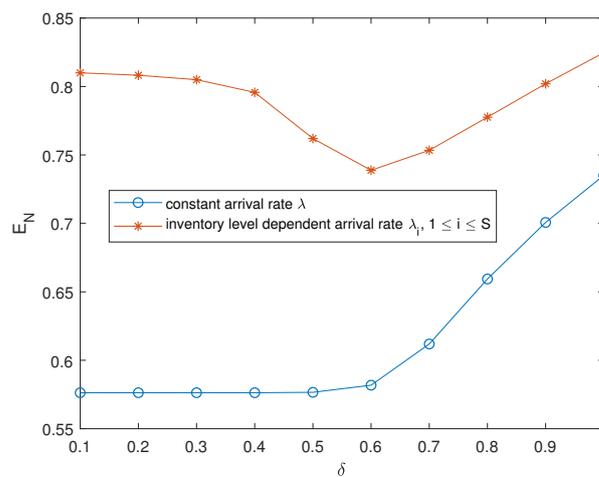


Figure 9. Effect of δ on E_N .

Interestingly, an increase in the value of δ initially results in a decrease in the expected number of customers and then starts increasing with an increase in the value of δ in the non-homogeneous arrival case. However, in the homogeneous arrival case, as δ increases initially, the expected number of customers remains constant and then starts increasing.

These observations are dependent on the input values of the parameters.

5. Conclusions

We studied a multi-server production inventory model that uses emergency replenishment when a delay in the production process leads to an imminent stock out. Customers arrive according to a non-homogeneous Poisson process. The service time duration at all servers is independent with identically distributed exponential random variables. A customer receives one item from the inventory at the end of their service with the probability δ . When the on-hand inventory reaches zero, a local purchase (emergency replenishment) to bring the inventory level to 1 is made. Due to this, the inventory level never reaches zero. Under the stability condition, we established a long-run system state distribution. A cost function involving these control variables was constructed, and we numerically investigated the optimal values. The effects of various parameters on the system performance measures were also evaluated.

In a follow-up paper, we will extend the present work to the case of Markovian arrival process *MAP* and phase-type *PH* service time.

Author Contributions: Conceptualization, A.K., D.S. and A.Z.M.; methodology, A.K. and D.S.; software, D.S.; formal analysis, A.K. and D.S.; investigation, D.S.; writing original draft preparation, A.K. and D.S.; writing review and editing, A.K., D.S., A.Z.M. and J.S.; supervision A.K.; project administration J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors are grateful to the reviewers for their valuable comments and suggestions that improved the quality and the presentation of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Neuts, M.F. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*; The Johns Hopkins University Press: Baltimore, MD, USA, 1981.
2. Melikov, A.Z.; Molchanov, A.A. Stock optimization in transportation storage systems. *Cybern. Syst. Anal.* **1992**, *28*, 484–487. [[CrossRef](#)]
3. Sigman, K.; Simchi-Levi, D. Light traffic heuristic for an M/G/1 queue with limited inventory. *Ann. Oper. Res.* **1992**, *40*, 371–380. [[CrossRef](#)]
4. Berman, O.; Kaplan, E.H.; Shimshak, D.G. Deterministic approximations for inventory management at service facilities. *IIE Trans.* **1993**, *25*, 98–104. [[CrossRef](#)]
5. Berman, O.; Kim, E. Stochastic models for inventory managements at service facilities. *Commun. Statist. Stoch. Model.* **1999**, *15*, 695–718. [[CrossRef](#)]
6. Berman, O.; Sapna, K.P. Optimal service rates of a service facility with perishable inventory items. *Nav. Res. Logist.* **2002**, *49*, 464–482. [[CrossRef](#)]
7. Schwarz, M.; Sauer, C.; Daduna, H.; Kulik, R.; Szekli, R. M/M/1 queueing systems with inventory. *Queueing Syst.* **2006**, *54*, 55–78. [[CrossRef](#)]
8. Saffari, M.; Asmussen, S.; Haji, R. The M/M/1 queue with inventory, lost sale and general lead times. *Queueing Syst.* **2013**, *75*, 65–77. [[CrossRef](#)]
9. Baek, J.W.; Moon, S.K. The M/M/1 queue with a production-inventory system and lost sales. *Appl. Math. Comput.* **2014**, *233*, 534–544. [[CrossRef](#)]
10. Krishnamoorthy, A.; Raju, N. N -policy for (s, S) perishable inventory system with positive lead time. *Korean J. Comput. Appl. Math.* **1998**, *5*, 253–261.
11. Yue, D.; Zhao, G.; Yue, W. Analysis of a Multi-server Queueing-inventory System with Non-homogeneous Poisson Arrivals. In Proceedings of the QTNA'16: 11th International Conference on Queueing Theory and Network Applications, Wellington, New Zealand, 13–15 December 2016; pp. 1–5. [[CrossRef](#)]
12. Houghton, M.; Isotupa, K.P.S. A Continuous Review Inventory System with Lost Sales and Emergency Orders. *Am. J. Oper. Res.* **2018**, *8*, 343–359. [[CrossRef](#)]
13. Yue, D.; Qin, Y. A Production Inventory System with Service Time and Production Vacations. *J. Syst. Sci. Syst. Eng.* **2019**, *28*, 168–180. [[CrossRef](#)]
14. Krishnamoorthy, A.; Shajin, D.; Narayanan, V.C. Inventory with positive service time: A survey. In *Advanced Trends in Queueing Theory*; Anisimov, V., Limnios, N., Eds.; Series of Books “Mathematics and Statistics”, Sciences; ISTE & Wiley: London, UK, 2021; Volume 2, pp. 201–238.
15. Alfares, H.K. Inventory model with stock-level dependent demand rate and variable holding cost. *Int. J. Prod. Econ.* **2007**, *108*, 256–265. [[CrossRef](#)]