Journal of the Korean Statistical Society 45 (2016) 14–24

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

A production–inventory system with a Markovian service queue and lost sales

Jung Woo Baek [a], Seung Ki Moon [b,*]

[a] Department of Industrial Engineering, Chosun University, Gwangju, Republic of Korea

[b] School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 20 March 2015
Accepted 12 May 2015
Available online 6 June 2015

AMS 2000 subject classifications:
primary 60K25
secondary 60K20

Keywords:
$M/M/c$ queue
(s, S) production–inventory model
Lost sales
Regenerative process

ABSTRACT

We study an (s, S) production–inventory system with an attached Markovian service queue. A production facility gradually replenishes items in the inventory based on the (s, S) scheme, and the production process is assumed to be a Poisson process. In addition to the production–inventory system, c servers process customers that arrive in the system according to the Poisson process. The service times are assumed to be independent and identically distributed exponential random variables. The customers leave the system with exactly one item at the service completion epochs. If an item is unavailable, the customers cannot be served and must wait in the system. During this out-of-stock period, all newly arriving customers are lost. A regenerative process is used to analyze the proposed model. We show that the queue size and inventory level processes are independent in steady-state, and we derive an explicit stationary joint probability in product form. Probabilistic interpretations are presented for the inventory process. Finally, using mean performance measures, we develop cost models and show numerical examples.

1. Introduction

We study an (s, S) production–inventory system with an attached $M/M/c/\infty$ service queue. A production facility gradually replenishes items in the inventory based on the (s, S) scheme, and the production process is assumed to be a Poisson process. In the production–inventory system, c servers are assigned to serve customers that arrive in the system according to the Poisson process. The service times are assumed to be independent and identically distributed (i.i.d.) exponential random variables. The customers leave the system with exactly one item at the service completion epochs. If the inventory level drops to zero, the remaining customers must wait in the system to be served. During this out-of-stock period, all arriving customers are lost. Inventory models with an attached service queue originated from an assembly-like queue (or kitting queue) (Bozer & McGinnis, 1992; Bryznér & Johansson, 1995; Harrison, 1973; Lipper & Sengupta, 1986), where several types of parts are simultaneously combined, and a positive processing time is incurred only after all the types are gathered. Sigman and Simchi-Levi (1992) conducted the first extensive study on an inventory model with an attached service queue. Using a light traffic heuristic for an $M/G/1$ queue with limited inventory, they provided a closed-form expression for the average delay in terms of basic system parameters. Even though previous studies presented various mean performance measures for inventory models with an attached service queue, research on the joint probability of the queue length and

* Corresponding author.
E-mail addresses: jwbaek@chosun.ac.kr (J.W. Baek), skmoon@ntu.edu.sg (S.K. Moon).

http://dx.doi.org/10.1016/j.jkss.2015.05.002
1226-3192/

inventory level was rare, even in simple cases, because of the analytic difficulty due to strict dependency between the queue length and inventory processes.

Schwarz and Daduna (2006) and Schwarz, Sauer, Daduna, Kulik, and Szekli (2006) conducted extensive research on the joint probability of the $M/M/1$ service queue with an attached inventory controlled by various policies with exponential lead time and backordering. An inventory model with a service queue and lost sales was also studied by Saffari, Haji, and Hassanzadeh (2011). The previous models derived the stationary joint probabilities of the queue length and inventory level in product form, which implied that the inventory process is independent of the queue length process in steady-state. Recently, the (s, S) production–inventory model with an $M/M/1$ service queue was studied by Krishnamoorthy and Viswanath (2013) where the inventory items were gradually replenished by an internal production process instead of being instantly supplied with the lead time. In this paper, the inventory model with single server queue, proposed by Krishnamoorthy and Viswanath (2013), is extended to the multi-server case. The purposes of our research are to obtain an explicit stationary joint probability in product form using a regenerative method and to provide the probabilistic interpretations of the probability.

Krishnamoorthy and Viswanath (2013) applied a matrix theoretic approach to derive the stationary joint distribution to efficiently and computationally analyze the system. Using this approach, they obtained comprehensive performance measures and a cost model; however, they showed no probabilistic interpretation of the inventory process. Saffari, Asmussen, and Haji (2013) employed a regenerative process to analyze an inventory queue with a service queue to facilitate more stochastic interpretations. They considered an (r, Q) inventory model with an attached $M/M/1$ service queue, general lead times, and lost sales. Deviating from Krishnamoorthy and Viswanath (2013), they proved that the inventory level process was independent of the queue length process in steady-state and derived the stationary joint probability. It is important to note that the method used by Saffari et al. (2013) ultimately led to separate analyses of the inventory and queue length processes. Using a similar methodology, Baek and Moon (2014) introduced an extended model called the (r, Q) production–inventory system with an attached service queue. In this proposed model, items were assumed to be stocked by both an outside supplier and internal production.

In this paper, we use a regenerative process for the analysis. Our proposed method allows for more probabilistic interpretations of the proposed model. First, we prove the independence between the queue length process and the inventory level process of the proposed model. Later, the queue length process and inventory level process are separately analyzed. For the analysis of the queue length and inventory level processes, the traditional $M/M/c/\infty$ and $M/M/1/K$ queues are applied to derive the stationary probabilities. Furthermore, we discuss the applicability of the proposed method to other production–inventory models with a Markovian service queue and lost sales. Finally, we conclude that the steady-state inventory level process becomes identical to the process described by Krishnamoorthy and Viswanath (2013), and we determine the optimal conditions of the proposed model for each of the decision variables s and S.

The remainder of this paper is organized as follows. In Section 2, we introduce the proposed model in more detail and show the preliminary results. In Section 3, we analyze the proposed model, and in Section 4, we present a cost model and optimal conditions for selecting decision values s and S. We also compare the proposed cost function with the cost function of Krishnamoorthy and Viswanath (2013). Finally, conclusions are discussed in Section 5.

2. Preliminaries

In this section, we introduce the proposed model and review the busy period queue length formula for the $M/M/1/K$ system, which plays an important role in the analysis.

2.1. The model

We study an $M/M/c/\infty$ service queue with an attached production–inventory system and lost sales as shown in Fig. 1. There are c servers dedicated to serving customers one by one under the first-come, first-served (FCFS) discipline. The waiting room for customers is unlimited, and the size of the inventory is S. Customers arrive in the system according to a Poisson process with rate λ. During the period when there are items, arriving customers join the queue; however, all customers that arrive during an out-of-stock period are lost. The service times are assumed to be i.i.d. exponential random variables with mean $1/\mu$, and we assume $\lambda < c\mu$ for the ergodicity of the queue. A customer leaves the system with exactly one item from the inventory at his service completion epoch. When the inventory level reaches zero, the remaining customers in the system wait until the inventory is replenished.

Fig. 2 shows the sample path of the proposed model when $S = 6$ and $s = 4$. Let $N(t)$ and $J(t)$ be the number of customers and inventory level at time t, respectively.

The inventory items are gradually replenished by an internal production facility that is controlled by the (s, S) policy; specifically, the production facility is turned off as soon as the inventory level becomes S and reactivated if the level drops to s. In this way, the system alternates between a production period and a non-production period; the lengths of these periods are denoted by L_P and L_N, respectively. We assume that the internal production process follows a Poisson process with rate η. For analytic convenience, we assume that $\eta \neq \lambda$. In the lower part of Fig. 2, it is easy to see that L_N is the first passage time to level s from level S, and L_P is the first passage time to level S from level s.

Note that the queue length and the on-hand inventory processes are completely dependent; however, the inventory level process behaves like a regenerative process, where all of the non-production period starting points are the regeneration

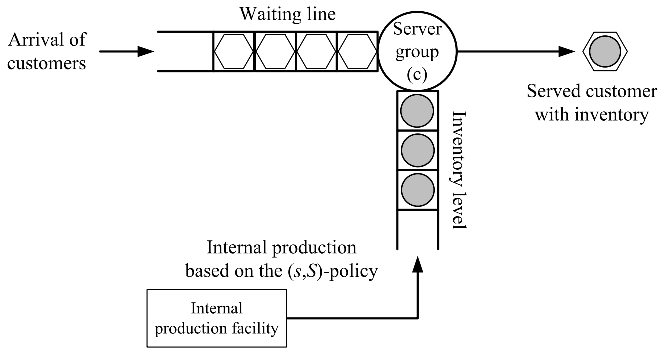


Fig. 1. $M/M/c/\infty$ queue with the production–inventory system.

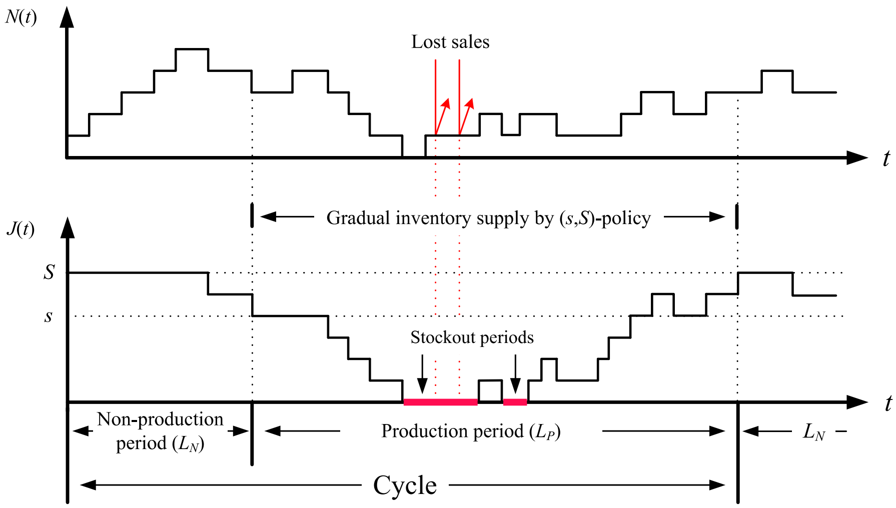


Fig. 2. Sample path of the proposed model.

points. This concept of the regenerative process plays an important role in obtaining the stationary joint probability of the number of customers and the on-hand inventory level and its stochastic interpretation. Later, we will show that the queue length process and inventory level process are independent in steady-state. The length from a starting point of a non-production period to the end of an ensuing production period is a cycle denoted by C. It follows that $C = L_N + L_P$.

2.2. The busy period queue length formula for the $M/M/1/K$ queue

The inventory level process during a production period is reminiscent of a finite capacity queueing system. In fact, in Section 3, we will prove that the inventory level process is stochastically identical to the conventional $M/M/1/K$ queue with s initial customers and upper absorbing boundary S. Therefore, the busy period queue length probability of the traditional $M/M/1/K$ queue plays an important role in our analysis. This section is based on results from Srivastava and Kashyap (1982), Takács (1962), and Takagi and Tarabia (2009).

Consider an $M/M/1/K$ queue with arrival rate λ, service rate μ, and system capacity K. Let $N^*(t)$ be the number of customers at time t. We define the following probability:

$$p_{(i)}(k, t|\lambda, \mu, K) = Pr\{N^*(t) = k, 0 < N^*(t) \le K \text{ for } 0 \le u \le t | N^*(0) = i\}, \quad 1 \le k \le K,$$

and the following Laplace transform (LT) and its generating function (GF):

$$P^*_{(i)}(k, \theta|\lambda, \mu, K) = \int_{t=0}^{\infty} e^{-\theta t} \cdot p_{(i)}(k, t|\lambda, \mu, K)dt, \qquad \tilde{P}^*_{(i)}(z, \theta|\lambda, \mu) = \sum_{k=1}^{S} z^k \cdot P^*_{(i)}(k, \theta|\lambda, \mu, K).$$

Using results from Srivastava and Kashyap (1982) and Takagi and Tarabia (2009), we have

$$\tilde{P}^*_{(i)}(z, \theta|\lambda, \mu, K) = \frac{z^{i+1} + \lambda z^{K+1}(1-z)P^*_{(i)}(K, \theta|\lambda, \mu, K) - z\mu P^*_{(i)}(1, \theta|\lambda, \mu, K)}{\theta z - (\mu - \lambda z)(1-z)}, \tag{1}$$

where

$$P^*_{(i)}(K,\theta|\lambda,\mu,K) = \frac{\left[\beta^*(\theta|\lambda,\mu)\right]^i - \left[\alpha^*(\theta|\lambda,\mu)\right]^i}{\lambda\cdot\left\{\left[\beta^*(\theta|\lambda,\mu)\right]^K\cdot\left[\beta^*(\theta|\lambda,\mu)-1\right]+\left[\alpha^*(\theta|\lambda,\mu)\right]^K\cdot\left[1-\alpha^*(\theta|\lambda,\mu)\right]\right\}}, \quad (2)$$

$$P^*_{(i)}(1,\theta|\lambda,\mu,K) = \left(\frac{\lambda}{\mu}\right)^{-i}\cdot\frac{\left[\beta^*(\theta|\lambda,\mu)\right]^{K-i}\cdot\left[\beta^*(\theta|\lambda,\mu)-1\right]+\left[\alpha^*(\theta|\lambda,\mu)\right]^{K-i}\cdot\left[1-\alpha^*(\theta|\lambda,\mu)\right]}{\mu\cdot\left\{\left[\beta^*(\theta|\lambda,\mu)\right]^K\cdot\left[\beta^*(\theta|\lambda,\mu)-1\right]+\left[\alpha^*(\theta|\lambda,\mu)\right]^K\cdot\left[1-\alpha^*(\theta|\lambda,\mu)\right]\right\}}, \quad (3)$$

and

$$\alpha^*(\theta|\lambda,\mu) = \frac{\theta+\lambda+\mu-\sqrt{(\theta+\lambda+\mu)^2-4\lambda\mu}}{2\lambda}, \quad (4)$$

$$\beta^*(\theta|\lambda,\mu) = \frac{\theta+\lambda+\mu+\sqrt{(\theta+\lambda+\mu)^2-4\lambda\mu}}{2\lambda}. \quad (5)$$

Remark 2.1. Note that (1) corrects the error in Takagi and Tarabia (2009).

Remark 2.2. Note that $p_{(i)}(k,t|\lambda,\mu,K)$ is the probability that k customers exist in the system at time t during a busy period. Therefore, $P^*_{(i)}(k,\theta|\lambda,\mu,K)|_{\theta=0} = \int_{t=0}^{\infty} p_{(i)}(k,t|\lambda,\mu,K)dt$ denotes the expected total sojourn time in level k during a busy period under the condition that the busy period starts with i initial customers. This expected sojourn time plays an important role in analyzing the production period.

2.3. Notation

Throughout this paper, we use the following notation:

S, s: decision variables, $(0 \le s \le S-1)$,
L_N: the length of a non-production period, (random variable),
L_P: the length of a production period, (random variable),
$C = L_N + L_P$: length of a cycle, i.e., the length of time between two successive non-production period starting points, (random variable),
$N(t)$: the number of customers in the queue at time t, $(N(t) \ge 0)$,
$J(t)$: the level of the on-hand inventory (number of items) at time t, $(0 \le J(t) \le S)$,
$N = \lim_{t\to\infty} N(t), J = \lim_{t\to\infty} J(t)$,
$P(n) = \lim_{t\to\infty} Pr[N(t)=n] = Pr(N=n)$, $(n \ge 0)$,
$\chi(k) = \lim_{t\to\infty} Pr[J(t)=k] = Pr(J=k)$, $(0 \le k \le S)$.

3. Analysis

In this section, we derive the joint distribution of the number of customers and inventory level at an arbitrary time in steady-state. First, we define the joint probability

$$\tilde{\Psi}(n,k,t|n_0,k_0) = Pr[N(t)=n, J(t)=k|N(0)=n_0, J(0)=k_0], \quad t \ge 0,\ n, n_0 \ge 0,\ \text{and } 0 \le k, k_0 \le S,$$

and the limiting probability,

$$\tilde{\Psi}(n,k) = \lim_{t\to\infty} \tilde{\Psi}(n,k,t|n_0,k_0), \quad (6)$$

where the effect of the initial conditions has faded away.

We begin the analysis with the proof of independence between $N(t)$ and $J(t)$ in steady-state. If these two random variables are independent, the stationary joint probability can be written as

$$\tilde{\Psi}(n,k) = P(n)\cdot\chi(k), \quad n,\ k \ge 0. \quad (7)$$

For the proposed queue with the production–inventory system described in Section 2, we assume an outside auxiliary inventory supplier provides one item instantaneously to the inventory as soon as the stock level becomes zero. This system with the auxiliary supply is called an $M/M/c/\infty$ queue with an (s,S) production–inventory system under the instant auxiliary supply.

Remark 3.1. To prevent confusion, and without loss of generality, we refer to the proposed model in Section 2 as the 'original system' and the above model with instant auxiliary supply as the 'modified system'.

Lemma 3.1. *In steady-state, the queue length process of the modified system is stochastically equivalent to the conventional $M/M/c/\infty$ queue. Moreover, the queue length process is independent of the on-hand inventory process.*

Proof. In the modified system, because of the instant auxiliary supply, no sales are lost, and there are no service interruptions. Therefore, the queue length process behaves identically to the traditional $M/M/c/\infty$ queue, which proves the equivalence between the two queue length processes.

Now, we note that the inventory level at time t only depends on the internal production process of the inventory items and the departure process of the queue during the time interval $(0, t]$. By assumption, the internal production process is independent of the queue length process. In addition, in steady-state, departure processes of customers in the modified system follow a Poisson process with rate λ, irrespective of the number of customers in the system; thus, the independence between queue length and inventory level processes is proved. □

Using Lemma 3.1, we then have the following theorem.

Theorem 3.1. *In steady-state, the queue length process of the original system, i.e., the proposed system in Section 2, is independent of the inventory level processes. In this case, the stationary joint probability $\tilde{\Psi}(n,k)$ of the original system is written as*

$$\tilde{\Psi}(n,k) = P_{M/M/c/\infty}(n) \cdot \chi(k), \quad n,\ k \geq 0, \tag{8}$$

where $P_{M/M/c/\infty}(n)$ is the stationary queue length probability of the conventional $M/M/c/\infty$ queue, which is given by

$$P_{M/M/c/\infty}(0) = \left[\left(\sum_{k=0}^{c-1} \frac{a^k}{k!} + \frac{a^c}{c!} \cdot \frac{1}{1-\rho}\right)\right]^{-1}, \quad \left(\rho = \frac{\lambda}{c\mu},\ a = \frac{\lambda}{\mu}\right), \tag{9}$$

and

$$P_{M/M/c/\infty}(n) = \begin{cases} \dfrac{(c\rho)^n}{n!} \cdot P_{M/M/c/\infty}(0), & (1 \leq n \leq c-1), \\ \dfrac{(c\rho)^c}{c!} \cdot \rho^{n-c} \cdot P_{M/M/c/\infty}(0), & (n \geq c). \end{cases} \tag{10}$$

Proof. Consider an 'excised system' of the original system where all the queue and inventory processes during out-of-stock periods are removed and the remainders are chronologically connected. In the excised system, one inventory item is instantaneously supplied as soon as the inventory level depletes to 0. Then, it is clear that the excised system is stochastically identical to the modified system in Lemma 3.1. Therefore, under the condition that the inventory level is positive, the stationary queue length distribution is independent of the inventory process and identical to the distribution of the traditional $M/M/c/\infty$ queue by Lemma 3.1.

Next, we investigate the stochastic behavior during out-of-stock periods. Since all out-of-stock periods are i.i.d. exponential random variables with mean $1/\eta$, we only need to find the queue length probability at a start of an out-of-stock period. In the original system, it is clear that the queue length at a start of an out-of-stock period is the same as the queue length at the end of the out-of-stock period. Here, the end of an out-of-stock period in the original system corresponds to a time point in the modified system at which one item is replenished by the auxiliary supply. Therefore, we can directly apply Lemma 3.1 to conclude that the stationary queue length probability at the start of an arbitrary out-of-stock period is independent of the inventory level and equivalent to the ordinary $M/M/c/\infty$ queue. This completes the proof. □

Remark 3.2. We note that the inventory level and the queue length processes stay frozen during the stock-out periods. This is why we can apply the 'modified system' for the proof of Theorem 3.1. Without the lost sales assumption, the methodology used in Theorem 3.1 is no longer available.

Remark 3.3. If $c = 1$ in (8)–(10), the same result obtained by Krishnamoorthy and Viswanath (2013) can be obtained. Moreover, taking $c \to \infty$, the result for an $M/M/\infty$ service queue can be obtained.

Remark 3.4. Theorem 3.1 is one of the main results in this paper. Note that we only use the output process of the service queue and the production process in the proof. Therefore, it is not difficult to infer that this approach can be applied to any type of Markovian service queue with an attached production–inventory system and lost sales as long as all of the out-of-stock periods are i.i.d. and the output process of the service queue follows a Poisson process under the condition that the inventory level is positive.

Remark 3.5. Because of the proof of independence between the queue and inventory level processes, we can directly use the stationary inventory level probability from previous research, if it exists. We will show in the following sections that the stationary inventory level process of the proposed model is stochastically identical to traditional (s, S) inventory model (without service queue) (Altiok, 1989). This is another significant implication of Theorem 3.1.

Using (8), the expected number $E(N)$ of customers in the system is given as

$$E(N) = \sum_{n=0}^{\infty}\sum_{k=0}^{S} n \cdot \tilde{\Psi}(n,k) = \sum_{n=0}^{\infty} n \cdot P_{M/M/c/\infty}(n) = \frac{(c\rho)^c}{c!} \cdot \frac{\rho}{(1-\rho)^2} \cdot P_{M/M/c/\infty}(0) + \frac{\lambda}{\mu}, \tag{11}$$

where $\rho = \lambda/(c \cdot \mu)$ and $P_{M/M/c/\infty}(0)$ are given in (9).

3.1. Analysis of the inventory level process

In this section, we derive the probability $\chi(k)$ of the inventory level at an arbitrary time in steady-state to completely determine $\tilde{\Psi}(n,k)$. For (s,S)-inventory system, Altiok (1989) derived the inventory level probability in a recursive form. The recursive solution is useful for a computational purpose. However, it is not efficient for probabilistic interpretation. Here, conditioning on the state of the production facility, we derive the probability $\chi(k)$ in a closed-form. We define the indicator function $\xi(t)$ as

$$\xi(t) = \begin{cases} 0, & \text{if a system is in a non-production period,} \\ 1, & \text{if a system is in a production period.} \end{cases}$$

We also define the following probabilities:

$$\chi_N(k,t) = Pr[J(t)=k, \xi(t)=0], \quad k=s+1, s+2, \ldots, S,$$
$$\chi_P(k,t) = Pr[J(t)=k, \xi(t)=1], \quad k=0,1,\ldots,S-1,$$

and their steady-state probabilities,

$$\chi_N(k) = \lim_{t\to\infty} \chi_N(k,t), \quad k=s+1, s+2, \ldots, S,$$
$$\chi_P(k) = \lim_{t\to\infty} \chi_P(k,t), \quad k=0,1,\ldots,S-1.$$

Thus, we obtain

$$\chi(k) = \chi_N(k) + \chi_P(k), \quad k=0,1,2,\ldots,S. \tag{12}$$

3.1.1. Analysis of the non-production period

In this section, we derive the probability $\chi_N(k)$ of the inventory level at an arbitrary time during a non-production period in steady-state. Note that the departure process of customers in the service queue is a Poisson process with rate λ in steady-state, if the inventory level is positive. Also, note that the output process of the inventory items is probabilistically identical to the departure process of the customers. Therefore, all of the state sojourn times in the non-production period are i.i.d. exponential random variables with mean $1/\lambda$ in steady-state. Since the non-production period starts with S inventory items and proceeds until the level becomes s, the expected length $E(L_N)$ of a non-production period is

$$E(L_N) = \frac{S-s}{\lambda}. \tag{13}$$

The above equation implies

$$\chi_N(k) = \frac{E(L_N)}{E(C)} \cdot \frac{\frac{1}{\lambda}}{E(L_N)} = \frac{1}{\lambda E(C)}, \quad (k=s+1, s+2, \ldots, S). \tag{14}$$

3.1.2. Analysis of the production period

In this section, we derive the probability $\chi_P(k)$ of the inventory level at an arbitrary time during a production period in steady-state. To find the probability, we use a dual inventory level process because the busy period queue length probability of the traditional $M/M/1/K$ queue can be used efficiently for the analysis. The dual inventory level process is defined by

$$\mathcal{J}^* = \{J^*(t), t\ge 0\} = \{S-J(t), t\ge 0\},$$

and the following probability and generating function (GF) are defined by

$$\chi_P^{dual}(k) = \lim_{t\to\infty} Pr[J^*(t)=k, \xi(t)=1], \quad k=1,\ldots,S,$$
$$\tilde{\chi}_P^{dual}(z) = \sum_{k=1}^{S} z^k \cdot \chi_P^{dual}(k).$$

Note that the dual process always starts with $(S-s)$ initial items because every production period starts with s inventory items. Also, if the inventory level is positive, the output process of the items is a Poisson process with rate λ in steady-state by Theorem 3.1. Thus, it is not difficult to show that the dual process is stochastically identical to the busy period of the ordinary $M/M/1/K$ queue with arrival rate λ, service rate η, $(S-s)$ initial customers, and system capacity S. Let $T_P^{dual}(k)$ be the total sojourn time in level k of the dual inventory process during a production period in steady-state. Eq. (1) and Remark 2.2 imply

$$\begin{aligned} E(L_p) &= \sum_{k=1}^{S} E\left[T_P^{dual}(k)\right] \\ &= \tilde{P}^{*}_{(S-s)}(z,\theta|\lambda,\eta,S)|_{z=1,\theta=0} = \sum_{k=1}^{S}\int_{t=0}^{\infty} p_{(S-s)}(k,t|\lambda,\eta,S)dt \\ &= \frac{S-s}{\eta-\lambda} - \frac{(\lambda/\eta)^{s+1}\cdot\left[1-(\lambda/\eta)^{S-s}\right]}{\eta\cdot\left[1-(\lambda/\eta)\right]^2}. \end{aligned} \tag{15}$$

Remark 3.6. Using results from Takagi and Tarabia (2009), it can be shown that $E(L_p)$ is equivalent to the expected busy period length of the ordinary $M/M/1/K$ queue starting with $(S-s)$ initial customers with system capacity S.

Eqs. (13) and (15) are used to obtain the following expected length $E(C)$ of a cycle:

$$E(C) = E(L_N) + E(L_P) = \frac{1}{\lambda\left(1-\frac{\lambda}{\eta}\right)}\left[S - s - \frac{\left(\frac{\lambda}{\eta}\right)^{s+2}\cdot\left[1-\left(\frac{\lambda}{\eta}\right)^{S-s}\right]}{\left(1-\frac{\lambda}{\eta}\right)}\right]. \tag{16}$$

We then have the following theorem.

Theorem 3.2. *Defining $T_P(k)$ as the total sojourn time in inventory level k during a production period of the original inventory process, it follows that $E\left[T_P(k)\right] = E\left[T_P^{dual}(S-k)\right]$ and*

$$\chi_P(k) = \frac{E\left[T_P(k)\right]}{E(C)} = \frac{E\left[T_P^{dual}(S-k)\right]}{E(C)} = \begin{cases} \dfrac{(\lambda/\eta)^{s-k} - (\lambda/\eta)^{S-k}}{E(C)\cdot(\eta-\lambda)}, & k = 0,1,\ldots,s, \\ \dfrac{1-(\lambda/\eta)^{S-k}}{E(C)\cdot(\eta-\lambda)}, & k = s+1, s+2\ldots, S-1. \end{cases} \tag{17}$$

Proof. The dual inventory process during a production period is stochastically identical to the busy period of the traditional $M/M/1/K$ queue with arrival rate λ, service rate η, $(S-s)$ initial customers, and system capacity S. It follows from (1)–(5) and Remark 2.2 that

$$\begin{aligned} \tilde{\chi}_P^{dual}(z) &= \frac{E(L_P)}{E(C)}\cdot\frac{P^{*}_{(S-s)}(k,\theta|\lambda,\eta,S)|_{\theta=0}}{E(L_P)} \\ &= \frac{z\cdot\left[z^{S-s} - 1 + \frac{\lambda(z-1)z^S\left[\left(\frac{\lambda}{\eta}\right)^s - \left(\frac{\lambda}{\eta}\right)^S\right]}{\eta-\lambda}\right]}{E(C)(1-z)(\eta-z\lambda)} \\ &= \frac{1}{E(C)(\eta-\lambda)}\left[\sum_{k=1}^{S-s} z^k - \sum_{k=1}^{S}\left(\frac{\lambda}{\eta}\cdot z\right)^k + z^{S-s}\cdot\sum_{k=1}^{s}\left(\frac{\lambda}{\eta}\cdot z\right)^k\right]. \end{aligned} \tag{18}$$

Using the definition of the GF $\tilde{\chi}_P^{dual}(z)$, we have

$$\chi_P^{dual}(k) = \frac{E(L_P)}{E(C)}\cdot\frac{E\left[T_P^{dual}(k)\right]}{E(L_P)} = \begin{cases} \dfrac{1-(\lambda/\eta)^k}{E(C)(\eta-\lambda)}, & k = 1,2,\ldots,S-s-1, \\ \dfrac{(\lambda/\eta)^{k-(S-s)} - (\lambda/\eta)^k}{E(C)(\eta-\lambda)}, & k = S-s, S-s+1,\ldots,S, \end{cases} \tag{19}$$

and the proof is complete. □

Then, using (14) and (17) in (12), we obtain the stationary probability $\chi(k)$ of the on-hand inventory level.

Remark 3.7. Using algebra, it can be shown that the stationary probability $\chi(k)$ of the inventory level is completely identical to the traditional (s,S) inventory model with lost sales (Altiok, 1989). Hence, we can infer that the stationary inventory process of the proposed model is stochastically identical to the traditional model with no service queue.

Using (12)–(14), (16), and (17), the expected inventory level $E(J)$ of the system is given by

$$E(J) = \sum_{n=0}^{\infty}\sum_{k=0}^{S} k \cdot \tilde{\Psi}(n, k) = \sum_{k=1}^{S} k \cdot \chi(k) = \frac{2 - \frac{(S-s)(S+s+3)}{\lambda E(C)}}{2\left(\frac{\lambda}{\eta} - 1\right)}. \tag{20}$$

4. Cost models and numerical examples

In this section, we develop a cost function with the machine reactivation level s and inventory size S. We consider a linear cost function $EC_1(s, S)$ with the following cost factors:

(a) Holding cost (C_{inv}): cost spent to keep an item in the system per unit time,
(b) Starting cost ($\tilde{K}$): fixed cost spent to turn on the production machine (a cost spent once in a cycle),
(c) Shortage cost (C_{loss}): cost incurred by the loss of a customer,
(d) Production cost (C_{rp}): cost of production per unit time per item,
(e) Waiting cost (C_{wait}): cost per unit time per customer incurred by a customer waiting during an out-of-stock period.

The average operating cost $EC_1(s, S)$ per unit time is given by

$$\begin{aligned} EC_1(s, S) &= C_{inv} \cdot E(J) + C_{rp} \cdot \eta \cdot \frac{E(L_p)}{E(C)} + C_{loss} \cdot \lambda \cdot \chi(0) + C_{wait} \cdot E(N) \cdot \chi(0) + \frac{\tilde{K}}{E(C)} \\ &= C_{inv} \cdot E(J) + C_{rp} \cdot \eta \cdot \frac{E(L_p)}{E(C)} + \lambda \cdot \left[C_{loss} + C_{wait} \cdot \frac{E(N)}{\lambda}\right] \cdot \chi(0) + \frac{\tilde{K}}{E(C)}. \end{aligned} \tag{21}$$

Eq. (21) is used to examine the effect of customer waiting time costs during an out-of-stock period (the fourth term $C_{wait} \cdot E(N) \cdot \chi(0)$ in the first equality). Hence, if $C_{wait} = 0$, the cost function $EC_1(s, S)$ becomes the same as the cost function of Krishnamoorthy and Viswanath in Krishnamoorthy and Viswanath (2013). Here, note that $C_{wait} \cdot E(N)$ plays the role of a constant coefficient in the cost function with respect to the decision variables s and S. Therefore, the proposed cost function has the same properties as the cost function of Krishnamoorthy and Viswanath (2013); thus, we can use all of the results and properties therein for our cost analysis. Readers are referenced to Krishnamoorthy and Viswanath (2013) for more details on the convexity nature of (s, S)-inventory system.

Let $\sigma = \frac{\lambda}{\eta}$ and $\tilde{C}^*_{loss} = \left[C_{loss} + C_{wait} \cdot \frac{E(N)}{\lambda}\right]$. Using Lemma 3.3.1 and 3.3.2 in Krishnamoorthy and Viswanath (2013), the optimal conditions that the optimal values S^* and s^* satisfy can be obtained.

(1) The optimal condition for S^*:

$$\begin{aligned} &\frac{\left(\sigma^{S^*+2}\log(\sigma) - \sigma + 1\right)}{\lambda(\sigma-1)^2 E(C)}\left[C_{inv}\frac{(S^*-s)(S^*+s+3)}{2(\sigma-1)} + \eta(S^*-s)(C_{rp} - \tilde{C}^*_{loss}) - \lambda\tilde{K}\right] \\ &- \left[C_{inv}\frac{(2S^*+3)}{2(\sigma-1)} + \eta(C_{rp} - \tilde{C}^*_{loss})\right] = 0. \end{aligned} \tag{22}$$

(2) The optimal condition for s^* for fixed S:

$$\begin{aligned} &\frac{\left(\sigma^{s^*+2}\log(\sigma) - \sigma + 1\right)}{\lambda(\sigma-1)^2 E(C)}\left[C_{inv}\frac{(S-s^*)(S+s^*+3)}{2(\sigma-1)} + \eta(S-s^*)(C_{rp} - \tilde{C}^*_{loss}) - \lambda\tilde{K}\right] \\ &- \left[C_{inv}\frac{(2s^*+3)}{2(\sigma-1)} + \eta(C_{rp} - \tilde{C}^*_{loss})\right] = 0. \end{aligned} \tag{23}$$

Since S^* and s^* are real numbers, nearby integer numbers can be used to find the optimal values and optimal cost.

For fixed s, (a) and (b) in Table 1 show the convex nature of the cost function $EC_1(s, S)$ with respect to S. To compare our results to those of Krishnamoorthy and Viswanath (2013), we use the same parameter settings. The expected costs with zero waiting cost ($C_{wait} = 0$) are obviously the same as those of Krishnamoorthy and Viswanath (2013) as shown in Table 1. Furthermore, the optimal inventory size is $S^* = 16$, irrespective of C_{wait}, because the waiting cost $C_{wait} \cdot E(N)$ plays the role of a constant in the cost function. In Table 2, (a) and (b) show the cost function $EC_1(s, S)$ with respect to s for fixed S. We obtain the same result as Krishnamoorthy and Viswanath (2013) when $C_{wait} = 0$ as shown in Table 2. Furthermore, Fig. 3 shows graphical representation of the convexity nature of cost function $EC_1(s, S)$ under various parameter settings.

To examine the effect of the number of servers, we consider another cost function defined by

$$EC_2(s, S, c) = EC_1(s, S) + c \cdot C_{Server}, \tag{24}$$

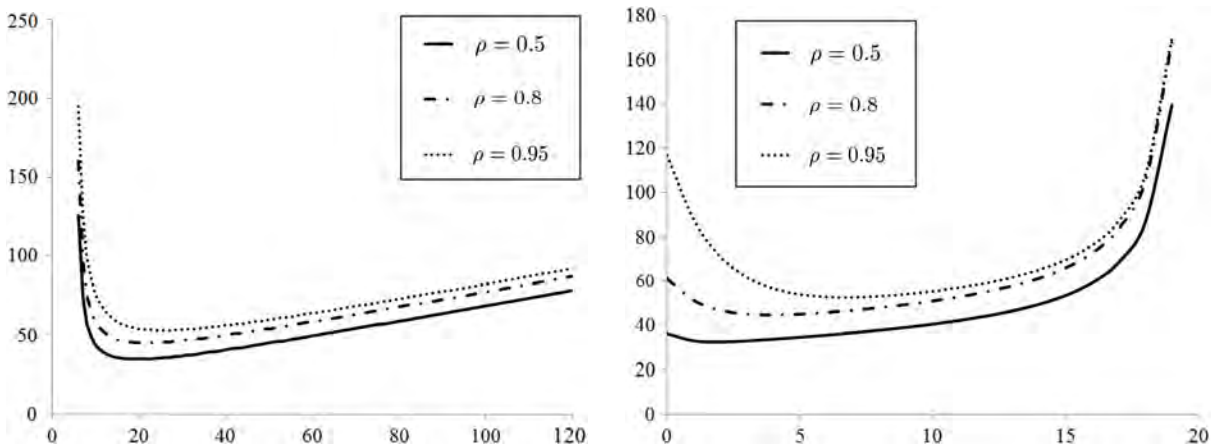
where C_{Server} is the per unit time per server cost for the service queue operation.

Table 1
Effect of the inventory storage size S on the cost function with respect to changes in C_{wait} and the fixed parameters $c = 1, s = 10, \lambda = 2, \mu = 3, C_{inv} = 50, C_{rp} = 200, C_{loss} = 400$, and $\tilde{K} = 2000$.

S	(a) $\eta = 3$			(b) $\eta = 1.5$		
	$C_{wait} = 0$	$C_{wait} = 100$	$C_{wait} = 200$	$C_{wait} = 0$	$C_{wait} = 100$	$C_{wait} = 200$
12	1249.17	1252.46	1255.76	653.954	705.349	756.745
13	1126.85	1129.81	1132.77	647.502	698.685	749.869
14	1077.09	1079.76	1082.43	645.182	696.178	747.175
15	1056.5	1058.92	1061.35	644.433	695.268	746.103
16	**1050.61**	**1052.82**	**1055.03**	**644.398**	**695.094**	**745.789**
17	1053.2	1055.22	1057.24	644.708	695.284	745.859
18	1061.14	1063	1064.85	645.178	695.652	746.126
19	1072.69	1074.4	1076.11	645.71	696.098	746.487
20	1086.8	1088.38	1089.96	646.248	696.565	746.882
21	1102.79	1104.26	1105.73	646.763	697.02	747.278
50	1759.68	1760.12	1760.55	649.994	699.994	749.994

Table 2
Effect of the machine reactivation level of inventory s on the cost function with respect to changes in c_w and the fixed parameters $c = 1, S = 15, \lambda = 2, \mu = 3, C_{inv} = 50, C_{rp} = 200, C_{loss} = 400$, and $\tilde{K} = 2000$.

s	(a) $\eta = 2.5$			(b) $\eta = 1.5$		
	$C_{wait} = 0$	$C_{wait} = 100$	$C_{wait} = 200$	$C_{wait} = 0$	$C_{wait} = 100$	$C_{wait} = 200$
2	827.3	836.0	844.7	639.7	691.5	743.2
3	842.4	849.7	857.0	639.9	691.5	743.1
4	861.1	867.2	873.3	640.3	691.7	743.2
5	883.0	888.1	893.3	640.7	692.0	743.4
6	908.1	912.5	916.9	641.2	692.5	743.7
7	936.9	940.7	944.4	641.8	693.0	744.1
8	970.1	973.3	976.5	642.6	693.6	744.6
9	1009.0	1011.8	1014.6	643.4	694.4	745.3
10	1056.5	1058.9	1061.4	644.4	695.3	746.1
12	1208.1	1210.0	1211.8	647.3	698.0	748.6



(a) Effect of S on $EC_1(s, S)$ when $s = 5$. (b) Effect of s on $EC_1(s, S)$ when $S = 20$.

Fig. 3. Expected costs $EC_1(s, S)$ with respect to changes in $\rho = \frac{\lambda}{\mu}$ under fixed parameters $\mu = 3, \eta = 5, C_{inv} = 1, C_{loss} = 200, C_{wait} = 100, \tilde{K} = 100, C_{rp} = 10$.

The following table shows the optimal number c^* that minimizes $EC_2(s, S, c)$ for given (s, S). For the cost $EC_2(s, S, c)$, we consider two cases, where the underlined values are the minimum costs for each case (see Table 3).

To determine the effect of the number of servers c on the expected cost $EC_2(s, S, c)$, we change the number of servers c under the condition that the other parameters are fixed. The results are shown in Fig. 4. For fixed s and S, $EC_1(s, S)$ is non-increasing as c increases because $E(N)$ is non-increasing with respect to c and converges to the expected number of customers for the $M/M/\infty$ queue, i.e., $E(N) = \frac{\lambda}{\mu-\lambda}$ when $c = 1$ and $\frac{\lambda}{\mu}$ when $c \to \infty$. On the other hand, the server operating cost $c \cdot C_{Server}$ increases as c increases. Therefore, $EC_2(s, S, c)$ has a convex nature with respect to c as shown in Fig. 4.

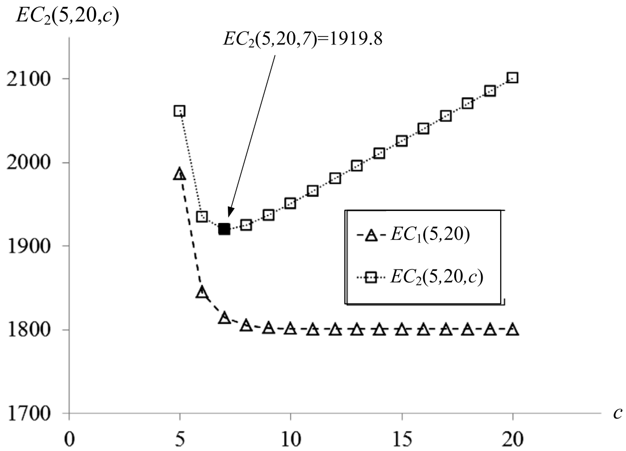


Fig. 4. Effect of the number of servers c on the expected cost under fixed parameters $s = 5$, $S = 20$, $\lambda = 12.5$, $\mu = 3$, $\eta = 5$, $C_{inv} = 1$, $C_{loss} = 200$, $C_{wait} = 100$, $\tilde{K} = 100$, $C_{rp} = 10$, and $C_{Server} = 15$.

Table 3
Expected costs $EC_2(s, S, c)$ with respect to changes in λ and c under fixed parameters $s = 10$, $S = 16$, $\mu = 3$, $C_{inv} = 50$, $C_{loss} = 400$, $C_{wait} = 200$, $\tilde{K} = 2000$, $C_{rp} = 200$, and $C_{Server} = 15$.

c	Case 1 ($\eta = 3$)				Case 2 ($\eta = 1.5$)			
	λ				λ			
	2.5	12.5	22.5	32.5	2.5	12.5	22.5	32.5
1	**872.539**	–	–	–	1189.71	–	–	–
2	874.669	–	–	–	**885.246**	–	–	–
3	889.176	–	–	–	888.011	–	–	–
4	904.114	–	–	–	901.467	–	–	–
5	919.106	5595.43	–	–	916.262	6060.88	–	–
6	934.105	5252.13	–	–	931.237	5661	–	–
7	949.105	5189.84	–	–	946.235	5586.51	–	–
8	964.105	**5181.03**	–	–	961.234	**5573.94**	–	–
9	979.105	5188.06	10284	–	976.234	5579.71	10713.8	–
10	994.105	5200.38	10017.1	–	991.234	5591.6	10425.3	–
11	1009.1	5214.5	9937.99	–	1006.23	5605.59	10338.9	–
12	1024.1	5229.23	9915.36	–	1021.23	5620.28	10313.4	–
13	1039.1	5244.15	**9914.4**	14959.1	1036.23	5635.19	**10311.2**	15376.4
14	1054.1	5259.13	9922.55	14754.5	1051.23	5650.16	10318.8	15160.6
15	1069.1	5274.12	9934.65	14678.2	1066.23	5665.15	10330.7	15079.6
16	1084.1	5289.12	9948.45	14650.9	1081.23	5680.15	10344.4	15050.3
17	1099.1	5304.12	9962.98	**14645.6**	1096.23	5695.15	10358.9	**15043.9**
18	1114.1	5319.12	9977.8	14650.7	1111.23	5710.15	10373.7	15048.5
19	1129.1	5334.12	9992.73	14660.9	1126.23	5725.15	10388.6	15058.4
20	1144.1	5349.12	10007.7	14673.6	1141.23	5740.15	10403.6	15071.1

5. Conclusion

In this paper, we studied (s, S) production–inventory systems with an attached $M/M/c/\infty$ service queue. We proved the independence between the inventory level and queue length processes and derived the explicit stationary joint probability in product form. Using the independence property, we separately analyzed the queue length probability and inventory level processes. For the inventory level probability, the concepts of the regenerative process and the traditional $M/M/1/K$ model were successfully applied, and we showed probabilistic interpretations for the proposed model using these results. We also discussed the applicability of our methodology to other production–inventory models with a Markovian service queue and lost sales. We showed the optimal conditions for each of the decision variables s and S. Furthermore, numerical examples were presented to find the optimal cost with respect to each of the decision variables.

We anticipate that the transient analysis of the proposed model will be completely different from the steady-state analysis because the queue length process is completely dependent on the inventory level process. In this situation, all the arguments used in this paper cannot be applied. Hence, the transient analysis of the proposed model is a topic of future research.

Currently, we assume inventory items are discrete. However, in real-world systems such as large-scale chemical plants, inventory may be continuous. Furthermore, in recently focused additive manufacturing systems and three-dimensional printing technologies, a product is made from uncountable materials such as polymer and metal powders. Therefore, a continuous inventory control model with an attached service queue is a challenging extension of the proposed model.

Acknowledgments

This work was supported by a start-up grant from Nanyang Technological University and an AcRF Tier 1 grant (RG94/13) from the Ministry of Education, Singapore.

References

Altiok, T. (1989). (R, r) production/inventory systems. *Operations Research*, *37*(2), 266–276.
Baek, J. W., & Moon, S. K. (2014). The M/M/1 queue with a production–inventory system and lost sales. *Applied Mathematics and Computation*, *233*(1), 534–544.
Bozer, Y., & McGinnis, L. (1992). Kitting versus line stocking: A conceptual framework and a descriptive model. *International Journal of Production Economics*, *28*(1), 1–19.
Bryznér, H., & Johansson, M. (1995). Design and performance of kitting and order picking systems. *International Journal of Production Economics*, *41*(1–3), 115–125.
Harrison, J. M. (1973). Assembly-like queues. *Journal of Applied Probability*, *10*(2), 354–367.
Krishnamoorthy, A., & Viswanath, N. C. (2013). Stochastic decomposition in production inventory with service time. *European Journal of Operational Research*, *228*(2), 358–366.
Lipper, E. H., & Sengupta, B. (1986). Assembly-like queues with finite capacity: Bounds, asymptotics and approximations. *Queueing Systems*, *1*(1), 67–83.
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.
Saffari, M., Haji, R., & Hassanzadeh, F. (2011). A queueing system with inventory and mixed exponentially distributed lead times. *The International Journal of Advanced Manufacturing Technology*, *53*(9–12), 1231–1237.
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.
Schwarz, M., Sauer, C., Daduna, H., Kulik, R., & Szekli, R. (2006). $M/M/1$ Queueing systems with inventory. *Queueing Systems*, *54*(1), 55–78.
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.
Srivastava, H. M., & Kashyap, B. R. K. (1982). *Special functions in queuing theory*. New York: Academic Press.
Takács, L. (1962). *Introduction to the theory of queues*. London: Oxford University Press.
Takagi, H., & Tarabia, A. M. K. (2009). Explicit probability density function for the length of a busy period in an $M/M/1/K$ queue. In *Advances in queueing theory and network applications* (pp. 213–226). New York: Springer.