An International Journal of Optimization and Control: Theories & Applications
ISSN: 2146-0957 eISSN: 2146-5703
Vol.12, No.1, pp.8-19 (2022)
https://doi.org/10.11121/ijocta.2022.1034

RESEARCH ARTICLE

Analysis of make-to-stock queues with general processing times and start-up and lost sales costs

Sinem Özkan*, Önder Bulut

Department of Industrial Engineering, Yaşar University, Turkey
sinem.ozkan@stu.yasar.edu.tr, onder.bulut@yasar.edu.tr

ARTICLE INFO

Article history:
Received: 19 October 2020
Accepted: 7 October 2021
Available Online: 23 October 2021

Keywords:
Production
Make-to-stock
Production and inventory control
Queueing theory
Renewal theory

AMS Classification 2010:
90B30, 90B05, 60K25, 60K05

ABSTRACT

We consider a make-to-stock environment with a single production unit that corresponds to a single machine or a line. Production and hence inventory are controlled by the *two-critical-number policy*. Production times are independent and identically distributed general random variables and demands are generated according to a stationary Poisson process. We model this production-inventory system as an M/G/1 make-to-stock queue. The main contribution of the study is to extend the control of make-to-stock literature by considering general production times, lost sales and fixed production costs at the same time. We characterize the long-run behaviour of the system and also propose a simple but very effective approximation to calculate the control parameters of the *two-critical-number policy*. An extensive numerical study exhibits the effects of the production time distribution and the system parameters on the policy control levels and average system cost.

1. Introduction

Most real-life production-inventory systems experience non-deterministic production and inter-demand-arrival times. In order to minimize the production and inventory related costs, performance evaluation and effective control of such systems are vital. This study addresses the production-inventory control problem of an environment with production start-up costs and general production times. Demands are generated according to a stationary Poisson process and the unsatisfied ones are immediately lost. The underlying system is controlled by the *two-critical-number policy* and modelled as an M/G/1 make-to-stock (MTS) queue. We develop a method to calculate the long-run average system cost. Furthermore, we determine the steady-state distribution of the inventory level and the production status (on or off) when production start-up cost is negligible.

The *two-critical-number policy* is known to be optimal for single-resource systems with backorders. Here, single-resource corresponds to a single machine or a single production line, which is the 'single-server' in queueing theory terminology. According to the *two-critical-number policy*, the production line is activated whenever inventory drops to the lower control level and production continues until the inventory reaches to the upper control level again.

To explain the practical significance of the problem, we can first scrutinize powder coating (powdered paint) production consisting of pre-mixing, extrusion and particle size reduction stages. Pre-mixing is the fastest stage of powder coating lines (can be further accelerated using additional caldrons) and therefore it can be assumed that extruder never starves. Although this process is a multi-stage one, in most real-life applications, it progresses in a continuous manner once the homogeneous mixture is obtained. After pre-mixing, without any interruption and intermediate buffers, particles are guided by air flow throughout the extrusion and size reduction stages. Thus, these two stages can be considered as a single operation while developing production policies. Furthermore, just before the start of a new production cycle the entire line is cleaned to get rid of dried paint chemicals and particle residues, which incur a start-up cost. In this paper, we aim to control production for single machine systems with ample supply and production start-up cost such as described above. Moreover, motivated from the powder coating example, it is also possible to generalize the practical use of our study to any lost sales make-to-stock system where the final station rarely

*Corresponding author

starves and production can be restarted with a fixed cost. Systems that are already single-station or have negligible start-up cost are special cases of this general perspective.

Our study extends the related literature by considering general production times, production start-up costs and lost-sales at the same time. Our contribution can be highlighted as: *i.* Majority of the earlier studies are for backordering environment. On the other hand, the studies with lost-sales either assume specific processing times (deterministic or Markovian) or zero start-up cost. We relax these restrictive assumptions and allow generally distributed processing times and non-zero start-up cost for the lost-sales case. *ii.* The first technical contribution is to develop an analogy between MTS queues and the capacitated M/G/1 queues. Using this analogy, the steady-state distribution of the considered MTS system is obtained when the production start-up cost is negligible. *iii.* For the systems with production start-up costs, we calculate the long-run expected average cost benefiting from renewal and queueing theories. We calculate the cycle cost with a first passage analysis and also show that the length of a production period can be written as the convolution of the lengths of capacitated M/G/1 busy periods. *iv.* We propose a well-performing approximation for the difference between the control levels of the *two-critical-number policy*.

The earlier studies consider single-server and single-demand-class systems with start-up and/or shut-down costs. The analyses are mostly based on queueing and inventory theory techniques. Gavish and Graves [1] is one of the initial milestone studies. They consider a single-demand-class backordering setting with production start-up cost. Production times are assumed to be constant and demands are generated according to a compound Poisson process with fixed sizes. For the two-critical-number policy, they calculate the average system cost and propose a search procedure to find the optimal values of the policy parameters. Gavish and Graves [2], and Lee and Srinivasan [3] extend [1] to general processing times. Both of these studies assume unit demand sizes. Graves and Keilson [4] and Srinivasan and Lee [5] consider compound demand extensions of [3]. Altıok [6] restricts production times to phase-type but studies both backordering and lost sales cases. Tijms [7] considers a system with general production times and finds the optimal control levels by a denumerable state Semi-Markov decision process. All the above-mentioned studies are actually MTS extensions of Heyman [8] and Sobel [9] who characterize the *two-critical-number policy* for the classical M/G/1 and G/G/1 queuing systems.

There are also related studies considering the systems where production occurs at a constant rate. De Kok et al. [10] analyses such a backordering production–inventory system where demands for a single product arrive according to a compound Poisson process and the production rate can be dynamically switched between two alternatives. Under some service level requirements, they derive approximations for the control parameters of the two-critical-number policy. De Kok [11] and De Kok and Tijms [12] study approximations for the lost-sales case. De Kok [13] provides an approximation for the time average of inventory holding and switching costs subject to a service level constraint. The recent work of Lin [14] considers a similar setting but the variation in the inventory level is modelled with a Brownian motion.

In more recent studies, the production-inventory problem is also considered as a control problem and Markov Decision Process (MDP) models are developed. The recent studies mostly assume negligible start-up costs. Ha [15] considers a setting with an exponential server, no start-up cost, several demand classes generating independent Poisson demands and lost-sales. He proves that the optimal production and rationing policies are of base-stock and static threshold level type, respectively. Bulut and Fadıloğlu [16] extends [15] and consider multiple production channels. They provide partial characterizations of the optimal policies and an extensive numerical study. Özkan and Bulut [17] considers the same environment with production start-up costs and proposes near optimal production and rationing policies.

Ha [18] considers the backordering version of [15] and characterizes the optimal policy by monotone switching curves for the two demand classes. For the same setting but with several demand classes, De Véricourt et al. [19] provides an algorithm to compute optimal rationing levels. Erlangian service times are considered by Ha [20] and Gayon et al. [21] for single-server lost-sales and backordering systems, respectively. They both show that threshold type policies based on work storage level are optimal. Pang et al. [22] allows batch demand and phase-type processing time distributions. Yücel and Bulut [23] assumes Coxian production times with non-zero production start-up cost. They propose an easy-to-apply, near optimal production policy.

In addition to the literature cited above, different system characteristics studied/assumed in the classical inventory control literature such as deteriorating items, more general demand structures, and environmental performance measures can also be adapted to production-inventory control literature. The following are example studies from the classical inventory and supply chain control literature that consider such characteristics: Pervin et al. [24], Tirkolaee et al. [25], Lofti et al. [26], and Paksoy et al. [27].

Our study is mostly related to [2] and [3]. We extend these studies to the lost-sales environment. The only existing study considering non-Exponential production times and lost sales is [6]. However, [6] is restricted to phase-type production times and thus Markovian analysis is still possible.

The rest of this paper is organized as follows: In Section 2, M/G/1 Make-to-Stock Queue with no production start-up cost is analysed and limiting probabilities and expected average system cost are calculated. In Section 3, we present a renewal approach to calculate the expected average system cost for the systems with production start-up costs. We also propose a near-optimal alternative for the difference of policy control levels. Section 4 is devoted to numerical experiments. Finally, Section 5 summarizes the study along with a discussion on future research directions.

2. Steady state analysis of M/G/1 make-to-stock queue with no start-up cost

We consider a make-to-stock (MTS) facility producing a single-item on a single production unit (e.g. a machine or a line). Demands are generated according to a stationary Poisson process with rate λ and production times are independent and identically distributed general random variables with rate μ. Unsatisfied demands are immediately lost and a cost of c is incurred per unit lost. Inventory holding cost rate is h.

In most of the production systems, inventory is replenished one-by-one at production completion instants. Hence, these systems are closely related to the classical queueing systems and are referred to as MTS queues. Our system is a lost-sales M/G/1 MTS queue that is controlled by the *two-critical-number policy*, in short (s, S) policy. S is the maximum inventory level at which production is stopped. The time period starting from S until reaching s is the non-production period. Whenever the inventory level drops to s, production is triggered with a start-up cost of K. The production-period continues until the inventory level reaches S again.

This section assumes negligible production start-up costs, i.e. $K = 0$. For such settings, Base-stock policy is optimal, i.e. $(s, S) = (S - 1, S)$. We conduct the steady-state analysis of the lost-sales M/G/1 MTS queue under $(S - 1, S)$ policy using the dynamics of the (typical) M/G/1/S queue. The analogy between two queues is given in Table 1. If there are $i \in \{0,1 \dots, S\}$ units of stock in the production-inventory system, then there are $(S - i)$ many "customers", which are the outstanding production orders, in the corresponding M/G/1/S queue.

Table 1. The analogy between 'M/G/1 MTS queue controlled by $(S - 1, S)$' and 'M/G/1/S queue'

	M/G/1 MTS Queue	M/G/1/S Queue
Customers	Demands	Outstanding production orders
System State	On-hand inventory	Outstanding production orders
Steady-state Probabilities $i = 0,1, \dots, S$	π_i^{MTSQ}	π_i^{Q}
	$\pi_i^{MTSQ} = \pi_{S-i}^{Q}$	

In order to calculate the long-run probabilities of the M/G/1/S queue, we follow the method proposed by Bose [28]. First, the steady-state distribution of the embedded Markov chain that tracks the system at the customer arrival instants is calculated. This distribution is equivalent to the steady state distribution of the considered M/G/1/S queue by PASTA (Poisson Arrivals See Time Averages) property. To obtain the steady state distribution of the embedded chain when the system is observed at the arrival instants, we benefit from the steady state distribution of another embedded chain that tracks the system at the departure instants.

The number of customers left behind in the M/G/1/S system at any customer departure instant in the long run follows a Markov Chain. The chain is described in (1). The state variable n_j is the number of customers left behind in the system at the time of j^{th} customer departure. The evolution of n_j depends on the number of new arrivals that occur during the service time of the j^{th} customer, which is denoted by a_j in (1), and the queue capacity $(S - 1)$.

$$n_{j+1} = \begin{cases} \min\{a_{j+1}, S-1\}, & n_j = 0, \\ \min\{n_j - 1 + a_{j+1}, S-1\}, & 1 \le n_j \le (S-1) \end{cases}. \quad (1)$$

The steady-state probabilities of the chain described in (1) are denoted by p_i^d where $i \in \{0,1 \dots, S-1\}$ is the number of customers left behind in the system at the time of any customer departure in the long run. That is,

$$p_i^d = \lim_{j\to\infty} P\{n_j = i\}. \quad (2)$$

These limiting probabilities can be easily calculated by using the transition probability matrix induced by (1). Once we characterize the steady state solution for the embedded Markov chain of M/G/1/S queue that tracks the system at the departure epochs, the next is to characterize the embedded chain that tracks the system at the arrival epochs. Due to Kleinrock's Result,

$$p_i^{ac} = p_i^d \quad (3)$$

where p_i^{ac}, $i \in \{0,1 \dots, S-1\}$, is the probability that an arrival (a new customer) finds i customers in the system in the long run. It should be noted that (3) holds only when the arrivals that find the system not full are accounted (Bose [28]). Therefore, the upper bound of i is $S - 1$.

The unconditional state probabilities at the arrival instants (regardless of whether the customer enters the system or leaves without joining the queue) are defined by $p_i^a, i \in \{0,1 \dots, S\}$. For all $i \in \{0,1 \dots, S-1\}$, the relation between the unconditional and conditional state probabilities (between p_i^a and p_i^{ac}) at the arrival instants is given in (4). Equation (4) also states that the steady state probability of the original system (π_i^Q) equals to the steady state probability when the system is observed at the arrival instants (p_i^a). This holds due to the PASTA property.

$$\pi_i^Q = p_i^a = \left(1 - \pi_S^Q\right) p_i^{ac} \tag{4}$$

In (4), $\left(1 - \pi_S^Q\right)$ is the probability that the system is not full in the long run and hence an arriving customer enters the system. In Equation (5), π_0^Q, which is the idleness rate of the system, is written in terms of π_S^Q.

$$\pi_0^Q = 1 - \rho = 1 - \frac{\lambda\left(1-\pi_S^Q\right)}{\mu} \tag{5}$$

In (5), ρ and $\lambda\left(1 - \pi_S^Q\right)$ denote the utilization and effective arrival rate of the system, respectively. Once the system of equations defined by (3), (4), and (5) is solved, the steady-state distribution of the M/G/1/S queue is obtained. Using the analogy given in Table 1, the steady state distribution of the M/G/1 MTS queue controlled by $(S-1, S)$ policy is found. Using the steady state distribution, the expected average system cost is calculated in (6).

$$\begin{aligned} AC(S-1,S) &= \lambda c \pi_S^Q + h \textstyle\sum_{i=1}^{S} i \pi_{S-i}^Q \\ &= \lambda c \pi_0^{MTSQ} + h \textstyle\sum_{i=1}^{S} i \pi_i^{MTSQ} \end{aligned} \tag{6}$$

The above analysis covers the cases where $s = (S-1)$. However, if a fixed start-up cost (K) is incurred to activate a line, then $s < (S-1)$ for most of the practical settings. For such instances of *the two-critical-number policy*, the adaptation of the described method is not direct. The main reason of the complication is the increase in the length of the non-production period, which can be measured by $\Delta = (S - s)$. If $\Delta > 1$, for any inventory level i such that $s < i < S$, it is not possible to certainly identify whether the server is on or off. Such states are visited both in the production and non-production periods. We overcome this extra level of complexity with a new method that directly calculates the long-run average cost without finding the steady-state distribution.

3. Expected long-run average cost of M/G/1 make-to-stock queue with start-up cost

In this section we consider the case where the production start-up cost K is positive. As [3] and [5] do for the backordering environment, we develop a renewal approach to calculate the average system cost.

We define the regeneration point of renewal cycles as the maximum inventory level S and decompose cycles into production and non-production periods (sub-cycles). When the inventory level reaches S, production is stopped and a new renewal cycle starts with its non-production period in which inventory is depleted by the demands. Upon hitting the lower control level s, non-production period ends and production period of the cycle starts. During this period, inventory level follows a realization increasing with production completions and decreasing with demand arrivals. Production period and the current renewal cycle is completed when the inventory level reaches S again. The basic notation used in the remaining part of the section is provided in Table 2.

Table 2. Basic notation of the renewal analysis

X	the random variable denoting the production time with cdf $G_X(x)$ and rate μ
$N(t)$	Poisson demand process with rate λ,
X_d	the random variable denoting the production time given that $N(X) = d$,
$E[X_d]$	$E[X \mid N(X) = d]$,
$C_N(s,S)$	expected cost of a non-production period,
$C_P(s,S)$	expected cost of a production period,
$L_N(s,S)$	expected length of a non-production period,
$L_P(s,S)$	expected length of a production period,
$AC(s,S)$	expected average cost.

By the Renewal Reward Theorem, the long-run expected average system cost can be written as follows:

$$AC(s,S) = \frac{C_N(s,S) + C_P(s,S) + K}{L_N(s,S) + L_P(s,S)} \tag{7}$$

Expected cost of a non-production period is relatively easy to calculate because the only relevant component is the holding cost and the average time spent at each inventory level $i \in \{s+1 \dots, S\}$ is $1/\lambda$. Then,

$$C_N(s,S) = \sum_{i=s+1}^{S} \frac{hi}{\lambda} \tag{8}$$

On the other hand, calculating the expected cost of a production period is more cumbersome. For $i, j \in \{0, \dots, S\}$ and $j \geq i$, if we let $E\left[\mathcal{F}_{i,j}\right] = f_{i,j}$ denote the expectation of the accumulated cost starting from the time instant when the process enters i until the inventory level is raised to j for the first time, then

$C_P(s,S)$ can be written as

$$C_P(s,S) = f_{s,S} = \sum_{i=s}^{S-1} f_{i,(i+1)} \quad (9)$$

In order to apply the stepwise approach defined by (9) we condition on the number of demand arrivals during the production time of an item.

$$f_{i,(i+1)} = E\left[\mathcal{F}_{i,(i+1)}\right] = E\left[E\left[\mathcal{F}_{i,(i+1)}|N(X)\right]\right] \quad (10)$$

The inner expectation of (10) can be calculated in two parts. In the first part, for each inventory level i and number of demand arrivals d, the cost accumulated throughout a production time period is calculated and denoted by P_{id}. The second part is to calculate the first passage cost of the new state, which is determined by the current state i and the number of demand arrivals d. Mathematically speaking, the inner expectation of (10) can be written as (11). In (11), given i and d, the new state is calculated as $(max(i-d,0)+1)$. The calculation is based on the lost sales assumption and the additional unit of inventory due to the production completion. For the sake of completeness, $f_{i,i}$ is set to zero for all inventory levels i.

$$E\left[\mathcal{F}_{i,(i+1)}|N(X)=d\right] = P_{id} + f_{(max(i-d,0)+1),(i+1)} \quad (11)$$

We calculate P_{id} in (12). The first part corresponds to the expected holding cost of all the items in the inventory. For each item $n \in \{1,\dots,i\}$, given that d demand arrivals occur during the production time, we let τ_{nd} be the length of the time that the n^{th} item spends in the inventory (holding time of the n^{th} item). In the second part, expected lost sales cost calculated where $(E[X_d]-\tau_{id})$ is the conditional expected length of the shortage period within the production time.

$$P_{id} = \sum_{n=1}^{i} h\tau_{nd} + c\,\lambda(E[X_d]-\tau_{id}) \quad (12)$$

To calculate τ_{nd} we make use of the following observation: holding time of the nth item is the arrival time of the nth demand. Since $N(t)$ is a Poisson process, the joint distribution of the conditional arrival times has the same distribution as the order statistics of independent Uniform random variables defined on the interval $[0,X_d]$ where X_d is the length of the production time given d demand arrivals, i.e., $N(X)=d$. Based on this fact, when $d>n$, the expected holding of the nth item is proportional to n divided by the total number of stochastically identical time intervals, which is $(d+1)$. When $d \le n$, on the other hand, nth item is held in the inventory during the whole production time. That is,

$$\tau_{nd} = \begin{cases} \frac{nE[X_d]}{d+1}, & d>n \\ E[X_d], & d \le n \end{cases} \quad (13)$$

Furthermore, using Bayes` Theorem and the shorthand notation $\alpha_d = P(N(X)=d)$, conditional expected production time $E[X_d]$ can be calculated as follows:

$$E[X_d] = E[X|N(X)=d]$$
$$= \frac{1}{\alpha_d}\int_u u\frac{e^{-\lambda u}(\lambda u)^d}{d!}\partial G(u) \quad (14)$$

where

$$\alpha_d = \int_u P(N(u)=d \mid U=u)\,\partial G(u)$$
$$= \int_u \frac{e^{-\lambda u}(\lambda u)^d}{d!}\partial G(u) \quad (15)$$

After obtaining the inner expectation, the outer expectation of (10), which is over the realizations of $N(X)$, can be written as,

$$f_{i,(i+1)} = E\left[P_{id} + f_{(\max(i-d,0)+1),(i+1)}\right]$$
$$= \sum_d \alpha_d\left(P_{id} + f_{(\max(i-d,0)+1),(i+1)}\right) \quad (16)$$

In order to find the values of the unknown $f_{i,j}$`s the system of linear equations defined by (16) can be solved recursively. The starting point is the calculation of $f_{0,1}$: $f_{0,1} = \sum_d \alpha_d\left(P_{id} + f_{1,1}\right) = \sum_d \alpha_d P_{id}$. Once the first passage costs are calculated, the expected cost of a production period can be compiled using (9).

After $C_N(s,S)$ and $C_P(s,S)$, the next step is to calculate the period lengths. Expected length of a non-production period can be directly written as

$$L_N(s,S) = \frac{S-s}{\lambda} \quad (17)$$

Although calculating production period length requires more effort, we succeed to develop a quick method that benefits from the analysis of capacitated M/G/1 queue. As it is discussed in Section 2, M/G/1 MTS queues are closely related with the capacitated M/G/1 queues. Our method to calculate $L_P(s,S)$ is based on the following observations:

(i) Production period starts when the inventory level drops to *s* from (*s*+1). This is equivalent to start a busy period of an M/G/1/(*s*+1) queue where the customers are the outstanding production orders. The capacity of the queue is (*s*+1) due to the lost sales assumption: there can be at most s more orders in addition to the one that triggers the busy period.

(ii) Within the busy period, each arrival event of the M/G/1/(*s*+1) queue decreases the inventory level of the original MTS system but each departure event on the other hand increases it. Furthermore, departure events are realized after the corresponding arrival events. Because of this zero-net flow, inventory level of the original MTS system at the end of the busy period would be (*s*+1), which is the inventory position when the busy period starts.

(iii) Similar to the above, once the inventory level hits (s+1), a busy period of M/G/1/(s+2) queue starts; and inductively this scheme repeats until the level reaches S.

If we let B_i denote the expected busy period length of an M/G/1/i queue, $i \in \{s+1 \dots, S\}$, then

$$L_P(s,S) = \sum_{i=s+1}^{S} B_i = \sum_{i=s+1}^{S} \left(\frac{1-\pi_0^{Qi}}{\lambda \pi_0^{Qi}}\right) \quad (18)$$

where π_0^{Qi} is the steady state probability of having zero customer in an M/G/1/i queue.

For given s and S values, all the components of (7), which are required to calculate the expected average system cost, are obtained. From the optimization perspective, on the other hand, we suggest the following EOQ-type formula for $\Delta = (S-s)$ instead of a two-dimensional enumerative search for the control parameters:

$$\Delta_{EOQ} = \left[\sqrt{\frac{2K\lambda}{h}}\right] \quad (19)$$

where $[.]$ returns the nearest integer.

Δ_{EOQ} is a near-optimal alternative for Δ^*, which is the optimal difference between the control parameters. Using Δ_{EOQ} we only search for s and then obtain S as $s + \Delta_{EOQ}$. In Section 4, it is shown that the proposed near-optimal solution performs very well.

4. Numerical study

In this section, we present the results of the numerical study and mainly show that how the optimal and near-optimal control levels and average cost react to changes in system parameters and processing time distributions. The discussion is enriched with managerial insights.

Study 1: We first show how changes in control levels s and S affect the average cost. For this study, the processing time is assumed to follow an Erlang-r distribution with parameters $\mu = 2$ and $r = 2$ where r is the number of processing stages (phases) and μ is the system processing rate (the rate of each stage is 4). That is, we consider an M/E$_2$/1 make-to-stock queue. Erlang distribution, which is a member of phase-type distribution family, enables to model production time distributions with different variance. While keeping μ constant, as r increases, variance of the production time decreases.

For the base case of the numerical study section, we let $(K, h, c, \lambda, \mu, r) = (10, 2, 40, 2, 2, 2)$. For the base case, Figure 1 exhibits that how the average cost changes with respect to s and S. The optimal control levels are $(s^*, S^*) = (5,9)$ and the optimal average cost is 15.66, which are found by enumerative search. Furthermore, for the base case, the approximation for the difference between control levels is calculated as Δ_{EOQ}= 4 that equals to the optimal difference $\Delta^*= (S^* - s^*) = 4$.

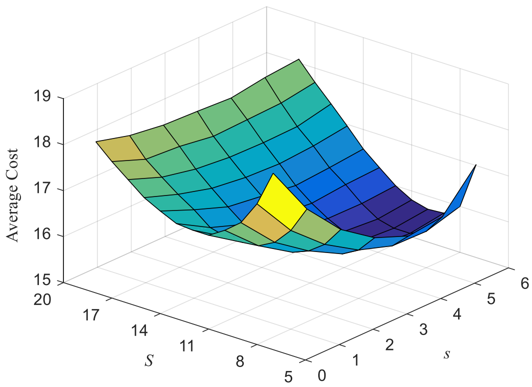

Figure 1. The average cost with respect to s and S

Study 2: In the second study, we aim to examine the impact of λ and $\Delta= (S-s)$ on the average cost. We present the study for $\lambda \in \{0.5, 1.0, 1.5, 2.0, 2.5\}$ and $\Delta \in \{1, 2, \dots, 15\}$. Given λ and Δ, we search for the optimal value of s and then calculate the optimal produce-up-to level as $S = \Delta + s$. We conduct the study for the base case introduced in Study 1 but this time traffic intensity of the system varies as λ changes.

In Figure 2, one can observe that when the traffic intensity $\rho = \frac{\lambda}{\mu}$ is lower, the average cost first decreases and then increases as Δ increases. On the other hand, as ρ increases, the flatness of the average cost function increases. That is, change in Δ does not affect the average cost significantly when the arrivals are more frequent. For such cases, due to heavy traffic, average inventory level would be low and shortage cost would be the dominant cost component.

In this study, we also would like to point out the use of constant Δ for the practitioners. When Δ is fixed, it is easy to calculate the expected length of a non-production period, which is Δ/λ. Such a fixation can be useful to solve the capacity allocation problem of multi-product systems. During the non-production period of an item, the available production capacity can be used for the others. Such a heuristic solution would simplify multi-product make-to-stock problems for the practitioners.

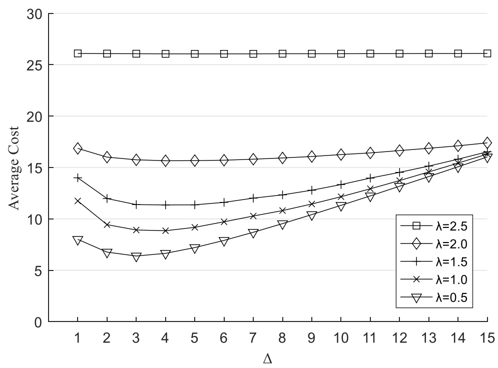

Figure 2. The impact of $\Delta= (S-s)$ on the average cost with different demand rates

Study 3: We now concentrate on the impact of various processing time distributions with the same mean ($E[X] = 0.50$) and variance ($\sigma^2(X) = 0.125$), and hence the same coefficient of variation ($c_X = \frac{\sigma(X)}{E[X]} = 0.71$). For this study, Log-normal, Erlang-r and Coxian-r distributions are considered. The first reason to select these distributions is that they have at least two parameters that enable us to work with the same mean and variance. Besides, the selected distributions are from different families; Conditional Normal and Phase-type. It is clear from Table 3 that as long as the first two moments of the processing time distributions are the same, the impact of different distributions on the average cost values and the optimality gaps is not remarkable. The optimality gap between $AC_{\Delta_{EOQ}}$ and AC^*, which are the costs of near-optimal and optimal solutions, are almost zero for all the cases. It can be concluded that our near optimal solution performs well under different processing time distributions having the same first two moments.

Study 4: We now focus on the impact of different $E[X]$ and $\sigma^2(X)$ values. For this purpose, we use the following parameters of the base case: $(K, h, c, \lambda) = (10, 2, 40, 2)$. The other parameters are displayed in Table 4 where $E[X]$ and $\sigma^2(X)$ are different for the considered distributions but c_X is the same for all.

In Table 4, distributions are placed from left to right in the descending order of means and variances. As one can expect that the average cost decreases as the first two moments of the processing time decrease. More interestingly, for all the considered cases, the optimality gap is almost zero. We can again conclude that EOQ-type formulation performs well under different processing time distributions and different distribution parameters.

Table 3. The impact of processing time distributions with same $E[X]$, $\sigma^2(X)$ and c_X

	Lognormal (0.50,0.354)	Erlang-2	Coxian-2 (3.92,3.92,0.96)
$E[X]$	0.50	0.50	0.50
$\sigma^2(X)$	0.125	0.125	0.125
c_X	0.71	0.71	0.71
AC^*	15.62	15.66	16.03
$AC_{\Delta_{EOQ}}$	15.62	15.66	16.05
Optimality Gap: $AC_{\Delta_{EOQ}}$ vs AC^*	0.00%	0.00%	0.12%

Table 4. The impact of processing time distributions with different $E[X]$ and $\sigma^2(X)$ but same c_X

	Lognormal (0.75,0.53)	Erlang-2	Coxian-2 (8,8,0.98)
$E[X]$	0.75	0.50	0.25
$\sigma^2(X)$	0.280	0.125	0.031
c_X	0.71	0.71	0.71
AC^*	29.70	15.66	11.54
$AC_{\Delta_{EOQ}}$	29.73	15.66	11.54
Optimality Gap: $AC_{\Delta_{EOQ}}$ vs AC^*	0.10%	0.00%	0.00%

Table 5. Processing time distributions with the corresponding parameter values

	Erlang-1	Erlang-2	Erlang-3	Erlang-5	Erlang-10	Erlang-500
$E[X]$	0.50	0.50	0.50	0.50	0.50	0.50
$\sigma^2(X)$	0.250	0.125	0.083	0.050	0.025	0.001
c_X	1.00	0.71	0.58	0.45	0.32	0.05

Table 6. The impact of processing time distributions

	Δ^*	s^*	AC^*	Δ_{EOQ}	$s_{\Delta_{EOQ}}$	$AC_{\Delta_{EOQ}}$	Optimality Gap: $AC_{\Delta_{EOQ}}$ vs AC^*
Erlang-1	5	5	17.41	4	6	17.43	0.11%
Erlang-2	4	5	15.66	4	5	15.66	0.00%
Erlang-3	4	5	14.96	4	5	14.96	0.00%
Erlang-5	4	5	14.32	4	5	14.32	0.00%
Erlang-10	4	4	13.84	4	4	13.84	0.00%
Erlang-500	4	4	13.31	4	4	13.31	0.00%

Study 5: This time we only fix the expected value of the processing time but let different variances. In order to assess the impact of such cases, we consider the base case of Study 1 and Erlang processing time distribution with different r (number of stages) values. While keeping the expected value constant, as r increases, the variance decreases and converges to zero in the limit (deterministic case).

Table 5 presents the considered processing time distributions with their means, variances and coefficient of variations. For the cases considered in Table 5, we provide optimal and near-optimal solutions, and optimality gaps in Table 6 where the distributions are ordered from high to low variances. While the variance decreases, Δ^*, s^*, and AC^* are non-increasing. As the processing time variance, s^* and Δ^* decrease, the lengths of both non-production and production periods also decrease. Therefore, the amount of cyclic safety stock and AC^* become less.

As it is seen in Table 6, Erlang-1, i.e. Exponential, distribution has the highest variance and therefore relatively higher optimality gap. On the other hand, optimality gaps are less than 0.12% for all the cases. Hence, it can be concluded that our EOQ-type formulation performs very well and is robust to changes in distribution and variance.

Study 6: Using the base case, in Figure 3, we investigate the joint effect of changes in production start-up cost K and Δ. When $\Delta > 10$, the cost function increases almost linearly and the effect of start-up cost on the average cost is limited. As Δ increases, the system pays less fixed cost and hence the portion of the start-up cost within the average cost decreases. As Δ increases, in the limit, the average cost would be same for all K values because the holding cost would dominate the other cost components. On the other hand, for smaller values of Δ, the effect of K is prominent: Lower start-up cost, lower the average cost.

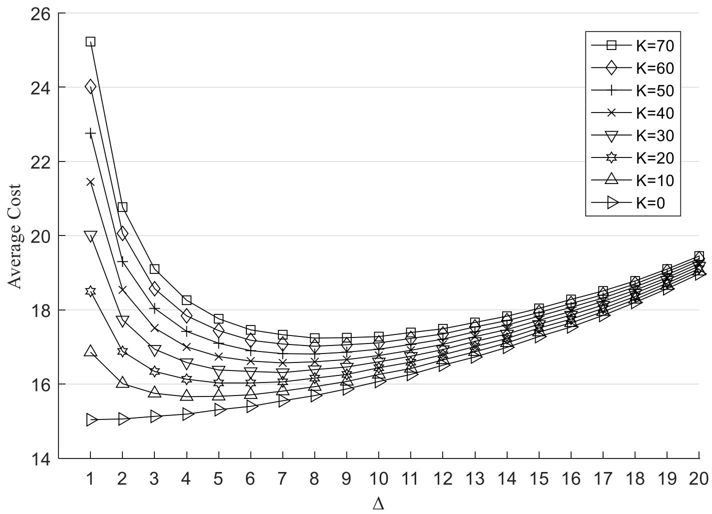

Figure 3. The average cost vs. $\Delta = (S - s)$ for each value of the fixed cost for Erlang-2

Study 7: We lastly examine the effects of changes in start-up cost K, lost-sales cost c, and holding cost h on optimal and near-optimal control levels, average costs, and optimality gaps. We conduct an extensive experimental study capturing different combinations of $K \in \{0,10,20\}$, $h \in \{1,2,3\}$, and $c \in \{h, 10h, 20h\}$. Tables 7 and 8 show the results for Erlang-2 and Uniform processing time distributions, respectively. For both of the distributions, expected processing time is 0.5.

As can be seen in Table 7, whatever h and c are, if the fixed cost to activate the line is negligible ($K = 0$), the optimal policy is a Base-Stock-type, i.e. $\Delta^* = (S^* - s^*) = 1$. For any given h and c values, as K increases, Δ^* increases to continue production for a longer time once it is triggered. On the other hand, s^* is non-increasing in K. Due to these behaviours, $S^* = s^* + \Delta^*$ is non-decreasing in K.

For a given K value, as c decreases or h increases, production is demotivated as inventory level increases. This is because of the trade-off between holding and shortage costs. Hence, the optimal production trigger point s^* and the optimal produce-up-to level S^*, diminish as c decreases or h increases. At the opposite direction, it is better to keep the line active at higher inventory levels to minimize the risk of stock out, which is now costlier.

The optimal average cost AC^* is increasing in K, c, and h as expected. However, the effect of K on the average cost is not as prominent as the effects of c and h. As K increases, both non-production and production period prolong and frequency of incurring the fixed cost reduces.

All the above observations and comments are also valid for the cases with Uniform processing time whose results are presented in Table 8.

To show the effectiveness of our heuristic approach, we compare the average cost calculated using Δ_{EOQ} with the optimized production trigger and produce-up-to levels, which are $s_{\Delta_{EOQ}}$ and $S_{\Delta_{EOQ}}$, respectively, with the optimal average cost. In Tables 7 and 8, the first one is represented as $AC_{\Delta_{EOQ}}$ and the second one as AC^*.

When $K = 0$, independent from the values of h and c are, the approximation finds the optimal solution. That is, for such cases the optimality gap is 0.00%. However, as K increases, optimality gap increases for small values of h and c. But, optimality gap is still around 0.00% for moderate and large values of h and c even for large K values. Considering all the cases, average optimality gaps are calculated as 0.38% and 0.55% for Erlang-r and Uniform distributions, respectively. In conclusion, we believe that our approach performs well at different levels of cost parameters under different distributions.

Table 7. Impact of system cost parameters for Erlang-2

K	h	c	Δ^*	s^*	S^*	AC^*	Δ_{EOQ}	$s_{\Delta_{EOQ}}$	$S_{\Delta_{EOQ}}$	$AC_{\Delta_{EOQ}}$	Optimality Gap: $AC_{\Delta_{EOQ}}$ vs AC^*
0	1	1	1	0	1	1.50	1	0	1	1.50	0.00%
0	1	10	1	4	5	5.25	1	4	5	5.25	0.00%
0	1	20	1	6	7	7.52	1	6	7	7.52	0.00%
0	2	2	1	0	1	3.00	1	0	1	3.00	0.00%
0	2	20	1	4	5	10.50	1	4	5	10.50	0.00%
0	2	40	1	6	7	15.04	1	6	7	15.04	0.00%
0	3	3	1	0	1	4.50	1	0	1	4.50	0.00%
0	3	30	1	4	5	15.75	1	4	5	15.75	0.00%
0	3	60	1	6	7	22.56	1	6	7	22.56	0.00%
10	1	1	5	0	5	3.08	6	0	6	3.12	1.18%
10	1	10	6	2	8	5.93	6	2	8	5.93	0.00%
10	1	20	5	5	10	8.01	6	4	10	8.01	0.00%
10	2	2	4	0	4	5.19	4	0	4	5.19	0.00%
10	2	20	4	3	7	11.37	4	3	7	11.37	0.00%
10	2	40	4	5	9	15.66	4	5	9	15.66	0.00%
10	3	3	4	0	4	7.15	4	0	4	7.15	0.00%
10	3	30	4	3	7	16.75	4	3	7	16.75	0.00%
10	3	60	4	5	9	23.25	4	5	9	23.25	0.00%
20	1	1	7	0	7	3.74	9	0	9	3.98	6.54%
20	1	10	7	2	9	6.29	9	1	10	6.35	0.88%
20	1	20	7	4	11	8.28	9	3	12	8.33	0.55%
20	2	2	5	0	5	6.17	6	0	6	6.24	1.18%
20	2	20	6	2	8	11.87	6	2	8	11.87	0.00%
20	2	40	5	5	10	16.03	6	4	10	16.03	0.02%
20	3	3	5	0	5	8.38	5	0	5	8.38	0.00%
20	3	30	5	3	8	17.35	5	3	8	17.35	0.00%
20	3	60	5	5	10	23.68	5	5	10	23.68	0.00%

Table 8. Impact of system cost parameters for Uniform (0.1,0.9)

K	h	c	Δ^*	s^*	S^*	AC^*	Δ_{EOQ}	$s_{\Delta_{EOQ}}$	$S_{\Delta_{EOQ}}$	$AC_{\Delta_{EOQ}}$	Optimality Gap: $AC_{\Delta_{EOQ}}$ vs AC^*
0	1	1	1	0	1	1.50	1	0	1	1.50	0.00%
0	1	10	1	4	5	4.86	1	4	5	4.86	0.00%
0	1	20	1	6	7	6.89	1	6	7	6.89	0.00%
0	2	2	1	0	1	3.00	1	0	1	3.00	0.00%
0	2	20	1	4	5	9.72	1	4	5	9.72	0.00%
0	2	40	1	6	7	13.78	1	6	7	13.78	0.00%
0	3	3	1	0	1	4.50	1	0	1	4.50	0.00%
0	3	30	1	4	5	14.58	1	4	5	14.58	0.00%
0	3	60	1	6	7	20.67	1	6	7	20.67	0.00%
10	1	1	5	0	5	2.96	6	0	6	3.02	1.88%
10	1	10	6	2	8	5.52	6	2	8	5.52	0.00%
10	1	20	6	4	10	7.37	6	4	10	7.37	0.00%
10	2	2	4	0	4	5.02	4	0	4	5.02	0.00%
10	2	20	4	3	7	10.55	4	3	7	10.55	0.00%
10	2	40	4	5	9	14.39	4	5	9	14.39	0.00%
10	3	3	4	0	4	6.95	4	0	4	6.95	0.00%
10	3	30	4	3	7	15.56	4	3	7	15.56	0.00%
10	3	60	4	5	9	21.38	4	5	9	21.38	0.00%
20	1	1	6	0	6	3.58	9	0	9	3.88	8.34%
20	1	10	7	2	9	5.86	9	1	10	5.95	1.54%
20	1	20	7	4	11	7.64	9	3	12	7.71	0.95%
20	2	2	5	0	5	5.93	6	0	6	6.04	1.88%
20	2	20	6	2	8	11.03	6	2	8	11.03	0.00%
20	2	40	6	4	10	14.74	6	4	10	14.74	0.00%
20	3	3	5	0	5	8.11	5	0	5	8.11	0.00%
20	3	30	4	3	7	16.10	5	2	7	16.13	0.17%
20	3	60	4	5	9	21.78	5	4	9	21.78	0.00%

5. Conclusion

We study a single-product make-to-stock system with a single production resource, production start-up cost, lost-sales, general production times and Poisson demand arrivals. Production and inventory are controlled by the *two-critical-number policy*.

As the production start-up cost is negligible, we find the steady-state distribution of the system with a method using the analogy between the considered MTS queue and the capacitated M/G/1 queue. In addition, for the systems with production start-up cost, we develop a method that directly calculates the long-run expected average cost. Our method benefits from renewal and queueing theories. We calculate the cycle cost with a first passage analysis. This analysis provides a system of equations, which is solved in a recursive manner. On the other hand, cycle length is calculated using busy period analysis of capacitated M/G/1 queues: We show that the length of a production period can be written as the convolution of the lengths of capacitated M/G/1 busy periods. Furthermore, we propose a well-performing approximation for the difference between the control levels of the *two-critical-number policy*. Finally, with an extensive numerical study, impacts of production time distributions, traffic intensity, and production start-up, lost-sales and holding costs are discussed.

As a possible extension of the study, multi-product or multi-demand class systems can be considered. A more compelling extension, on the other hand, would be examining multi-server systems.

References

[1] Gavish, B., Graves, S.C., (1980). A one-product production/inventory problem under continuous review policy. *Operations Research*, 28(5), 1228-1236.

[2] Gavish, B., Graves, S.C., (1981). Production/inventory systems with a stochastic production rate under a continuous review policy. *Computers & Operations Research*, 8(3), 169-183.

[3] Lee, H.S. Srinivasan, M.M., (1987). *The continuous review (s, S) policy for production/inventory systems*

with Poisson demands and arbitrary processing times. Technical Report, 87-33.

[4] Graves, S.C., Keilson, J., (1981). The compensation method applied to a one-product production/inventory problem. *Mathematics of Operations Research*, 6(2), 246-262.

[5] Srinivasan, M.M., Lee, H.S., (1991). Random review production/inventory systems with compound Poisson demands and arbitrary processing times. *Management Science*, 37(7), 813-833.

[6] Altiok, T., (1989). (R, r) production/inventory systems. *Operations Research*, 37(2), 266-276.

[7] Tijms, H.C., (1980). An algorithm for average costs denumerable state semi-Markov decision problems with application to controlled production and queueing system. In R. Hartley, L. C. Thomas, & D. J. White, Recent developments in Markov decision processes (pp. 143-179), Academic Press.

[8] Heyman, D. P., (1968). Optimal operating policies for M/G/1 queuing systems. *Operations Research*, 16(2), 362-382.

[9] Sobel, M.J., (1969). Optimal average-cost policy for a queue with start-up and shut-down costs. *Operations research*, 17(1), 145-162.

[10] De Kok, A.G., Tijms, H.C., Van der Duyn Schouten, F.A., (1984). Approximations for the single-product production-inventory problem with compound Poisson demand and service-level constraints. *Advances in Applied Probability*, 16(2), 378-401.

[11] De Kok, A.G., (1985). Approximations for a lost-sales production/inventory control model with service level constraints. *Management Science*, 31(6), 729-737.

[12] De Kok, A.G., Tijms, H.C., (1985). A stochastic production/inventory system with all-or-nothing demand and service measures. *Communications in Statistics. Stochastic Models*, 1(2), 171-190.

[13] De Kok, A.G., (1987). Approximations for operating characteristics in a production-inventory model with variable production rate. *European journal of operational research*, 29(3), 286-297.

[14] Lin, H.J., (2017). Two-critical-number control policy for a stochastic production inventory system with partial backlogging. *International Journal of Production Research*, 55(14), 4123-4135.

[15] Ha, A.Y., (1997a). Inventory rationing in a make-to-stock production system with several demand classes and lost sales. *Management Science*, 43(8), 1093-1103.

[16] Bulut, Ö., Fadiloğlu, M.M., (2011). Production control and stock rationing for a make-to-stock system with parallel production channels. *IIE Transactions*, 43(6), 432-450.

[17] Özkan, S., Bulut, Ö. (2020). Control of make-to-stock production systems with setup costs. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 35(3), 1199-1212.

[18] Ha, A.Y., (1997b). Stock-rationing policy for a make-to-stock production system with two priority classes and backordering. *Naval Research Logistics (NRL)*, 44(5), 457-472.

[19] De Véricourt, F., Karaesmen, F., Dallery, Y., (2002). Optimal stock allocation for a capacitated supply system. *Management Science*, 48(11), 1486-1501.

[20] Ha, A.Y., (2000). Stock rationing in an M/Ek/1 make-to-stock queue. *Management Science*, 46(1), 77-87.

[21] Gayon, J.P., De Vericourt, F., Karaesmen, F., (2009). Stock rationing in an M/Er/1 multi-class make-to-stock queue with backorders. *IIE Transactions*, 41(12), 1096-1109.

[22] Pang, Z., Shen, H., Cheng, T.C.E., (2014). Inventory Rationing in a Make-to-Stock System with Batch Production and Lost Sales. *Production and Operations Management*, 23(7), 1243-1257.

[23] Yücel, Ö., Bulut, Ö., (2019). Control of M/Cox-2/s make-to-stock systems. *An International Journal of Optimization and Control: Theories & Applications (IJOCTA)*, 10(1), 26-36.

[24] Pervin, M., Roy, S.K., Weber, G.W. (2018). Analysis of inventory control model with shortage under time-dependent demand and time-varying holding cost including stochastic deterioration. *Annals of Operations Research*, 260(1), 437-460.

[25] Tirkolaee, E.B., Goli, A., Weber, G.W. (2019). Multi-objective aggregate production planning model considering overtime and outsourcing options under fuzzy seasonal demand. In *International Scientific-Technical Conference Manufacturing*, (pp. 81-96). Springer, Cham.

[26] Lotfi, R., Weber, G.W., Sajadifar, S.M., Mardani, N. (2020). Interdependent demand in the two-period newsvendor problem. *Journal of Industrial & Management Optimization*, 16(1), 117-140.

[27] Paksoy, T., Özceylan, E., Weber, G.W. (2010). A multi objective model for optimization of a green supply chain network. In *AIP conference proceedings* (Vol. 1239, No. 1, pp. 311-320). American Institute of Physics.

[28] Bose, S. K. (2002). Basic Queueing Theory. In *An Introduction to Queueing Systems* (pp. 9-54). Springer, Boston, MA.

Sinem Özkan *received her MSc degree from the Industrial Engineering Department of Yaşar University in 2016. She has been continuing her PhD studies in the same department since 2016. Her research interests include production/inventory systems and stochastic modeling.*
http://orcid.org/0000-0002-7181-5800

Önder Bulut *is Assistant Professor in the Department of Industrial Engineering at Yaşar University. He received his Ph.D. degree from the Industrial Engineering Department of Bilkent University. His research interests include stochastic modeling, production and inventory systems, simulation, dynamic programming and optimal control.*
http://orcid.org/0000-0003-1476-6333

An International Journal of Optimization and Control: Theories & Applications (http://ijocta.balikesir.edu.tr)