5-12-2021

Queuing-Inventory Models with MAP Demands and Random Replenishment Opportunities

Srinivas R. Chakravarthy
Kettering University, schakrav@kettering.edu

B. Madhu Rao

mathematics

MDPI

Article

Queuing-Inventory Models with MAP Demands and Random Replenishment Opportunities

Srinivas R. Chakravarthy [1,*] and B. Madhu Rao [2]

[1] Departments of Industrial and Manufacturing Engineering & Mathematics, Kettering University, Flint, MI 48504, USA

[2] Department of Business Systems and Analytics, School of Business Administration, Stetson University, Deland, FL 32723, USA; bmrao@stetson.edu

* Correspondence: schakrav@kettering.edu; Tel.: +1-810-762-7906

Abstract: Combining the study of queuing with inventory is very common and such systems are referred to as queuing-inventory systems in the literature. These systems occur naturally in practice and have been studied extensively in the literature. The inventory systems considered in the literature generally include (s, S)-type. However, in this paper we look at opportunistic-type inventory replenishment in which there is an independent point process that is used to model events that are called opportunistic for replenishing inventory. When an opportunity (to replenish) occurs, a probabilistic rule that depends on the inventory level is used to determine whether to avail it or not. Assuming that the customers arrive according to a Markovian arrival process, the demands for inventory occur in batches of varying size, the demands require random service times that are modeled using a continuous-time phase-type distribution, and the point process for the opportunistic replenishment is a Poisson process, we apply matrix-analytic methods to study two of such models. In one of the models, the customers are lost when at arrivals there is no inventory and in the other model, the customers can enter into the system even if the inventory is zero but the server has to be busy at that moment. However, the customers are lost at arrivals when the server is idle with zero inventory or at service completion epochs that leave the inventory to be zero. Illustrative numerical examples are presented, and some possible future work is highlighted.

Keywords: queuing-inventory systems; algorithmic probability; batch demands; random opportunities; lead times; matrix-analytic methods

check for updates

Citation: Chakravarthy, S.R.; Rao, B.M. Queuing-Inventory Models with Random Replenishment Opportunities. *Mathematics* **2021**, *9*, 1092. https://doi.org/10.3390/math9101092

Academic Editor: Vladimir M. Vishnevsky

Received: 18 April 2021
Accepted: 10 May 2021
Published: 12 May 2021

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

1. Introduction

Models for inventory management under uncertainty have been studied extensively. The two key questions of when and how many to order have been addressed under a variety of factors such as the nature of inventory review (continuous or periodic), order quantity (fixed or variable), lead time for an order to be fulfilled (negligible, constant or random), nature of demand (deterministic or random), and other factors to optimize a function of various costs such as ordering, carrying inventory, lost sales, etc.

Most models assume a single supplier and fixed cost of replenishment. Some models incorporate the availability of random opportunities for replenishment which may lower system costs due to reduced unit cost and/or ordering cost. We refer to them as opportunistic replenishment. Friend [1] studied systems with special opportunities occurring according to a Poisson process. These opportunities, which may be exercised only at the instant of their occurrence, are offered for the same unit price but at reduced or zero ordering cost. Hurter and Kaminsky [2] extend this model to systems where the special opportunities may be exercised during a random period. Silver, Robb, and Rahnama [3] developed an efficient heuristic for the Hurter and Kaminsky [2] model. Gurnani [4] considered periodic review systems and used game theory to study the economics of placing a special order between reviews if a discounted sale is offered. Moinzadeh [5] considered systems with constant demand rate, zero lead time, and special discounted opportunities occurring at exponentially distributed intervals, and obtained optimal order quantity at regular price when inventory reaches zero, and the threshold and order quantity at discounted price. Feng and Sun [6] suggest modifying the (s, S) policy to a four-parameter system (threshold and order quantity for regular and discount purchases) and proposed a bisection search procedure to determine the optimal values of the four parameters. Goh and Sharafali [7] consider the model in [1,2] and incorporate the policy of passing the cost reduction due to special purchases downstream to increase demand. Chaouch [8] considers the model in [1,2] when special and regular prices are valid over alternating exponential intervals. Tajbakhsh and Zolfaghari [9] consider systems with discounts offered at exponential intervals with the discount price given by a discrete random variable and develop optimal order quantities at each special price and further extended the model to the case of multiple suppliers. Karimi-Nasab and Konstantaras [10] consider a system with constant demand, periodic random review intervals (uniformly distributed or exponential subject a maximum and minimum) and random special sale offers, and determine maximum inventory level for regular purchases, and a higher maximum for special purchases. Den Boer and Zwart [11] consider a system where the management makes simultaneous decisions on whether to take advantage of the special discounted offer, and the selling price of a unit similar to [7]. All the above references assume that the lead time for receiving a special replenishment is negligible.

In all the above models, it is assumed that the special opportunities for replenishment are always considered as a supplement to the normal replenishment process. In many situations such as drugstores, groceries, small supermarkets, etc., the suppliers visit the retailers at random (but frequent) intervals to offer special sales. This raises the possibility that for some systems it may be more economical to manage inventory solely based on special offers. This is particularly attractive for non-critical items (e.g., canned goods, generic medication) where stock outs are not critical. Special replenishment opportunities offer lower unit cost and/or reduced ordering cost, and are usually available without delay.

Inventory models discussed above assume that the time to process a demand is negligible. In many situations, completing a customer's service requires other resources, in addition to the item(s) in inventory (e.g., requiring a pharmacist to process the sale for prescription medication). Such inventory systems are referred to as queuing-inventory (QI) models or inventory models with positive service time, and have received a great deal of attention recently. Research in queuing-inventory systems may be classified based on features such as the nature of customer arrival process, service time distribution, server vacation, service discipline, customer behavior, inventory review policy, replenishment policy, stock-out assumptions, perishability of units, lead time for replenishment, among others. We refer the readers to Krishnamurthy, Shajin, and Narayanan [12] for a general review of queuing-inventory models, to Choi et al. [13] for a survey of QI models with phase-type service times, and to Melikova and Shahmaliyeva [14] for a discussion of QI models for perishable items. Recent generalizations of QI models are due to Chakravarthy, Maity and Gupta [15], who studied a QI model in which the service is carried out in batches, Chakravathy [16], who studied a QI model in which the demand occurs in batches, and Chakravarthy and Rumyantsev [17], who study a QI system with batch demand, Markovian Arrival Process (MAP), and phase-type service time and exponential replenishment lead times.

In this paper, we explore queuing-inventory systems which solely use special opportunities for replenishment. This research is at the interface of the two areas of traditional inventory systems, and the queuing-inventory systems. Most traditional inventory models consider the demand to occur one unit at a time, at either constant or exponential intervals. Our models extend this to consider demand that can occur in batches of random size, with arrivals following a very general process which allows for a broad class of inter-arrival times and autocorrelation between inter-arrival times. For both types of inventory systems, it is a significant departure to manage inventory exclusively based on random replenishment opportunities. The objective of this study was to understand the behavior of such systems and compare them with the traditional (s, S)-type inventory management systems.

2. General Description

The customer arrivals are modeled using Neuts' versatile Markovian point process, now referred to as batch Markovian arrival process ($BMAP$). If customers arrive one at a time, the process is referred to as Markovian arrival process or briefly MAP. While the customers arrive singly to the service facility, but their demands for inventory items are in varying sizes. One of the major advantages of using MAP to model the arrival process is to capture any correlation present in successive arrivals. The MAP is completely described by two parameter matrices, say, (E_0, E_1) of dimension m. While, E_0 governs transitions within the underlying Markov process with generator $E = E_0 + E_1$, the transitions governed by E_1 correspond to the arrivals of customers. We assume that E is an irreducible generator. MAP and $BMAP$ have been extensively used in stochastic model and there is a vast literature available on these. We refer the reader to [18–29] for details on MAP and $BMAP$ and other key references.

Each customer demands a random number of items from inventory before being processed by a single server. At the instant of an arrival, the available inventory is reduced by the amount demanded by the customer. The number of remaining items in inventory is referred to as available to promise or ATP in the business community. In this paper, we always refer to ATP when referring to inventory. Physical inventory is not relevant because once allocated to a customer, a unit is no longer available.

When the inventory level is positive, but less than the number of units demanded by a customer, the customer's demand is partially satisfied and the inventory level is set to zero. Arrivals to the system when inventory level is zero, are denied entry and are lost. The time to service a customer has a phase-type or PH-distribution and is independent of the number of units demanded. For details on PH-distributions and their usefulness in stochastic modeling, we refer the reader to [29–31].

Opportunities to replenish the inventory occur according to a Poisson process. We assume that the lead time associated with a replenishment is zero. Replenishment decisions are based on two parameters K and L ($0 \leq L < K < \infty$) and for convenience we refer to the policy as a (K, L) policy. If the inventory at the time a replenishment opportunity becomes available is equal to i, a replenishment order to bring the inventory level to K is placed with probability 1 if $i \leq L$, and with with probability a_{i-L} if $L < i < K$.

In this paper, we consider two types of models, referred to as Opportunistic Model 1 and Opportunistic Model 2, and these are described in the following sections.

Any queuing-inventory model with positive service time can be looked at two different ways according to these two models to see which one will be better from either customer or management or from both points of view. Therefore, there is a trade-off between the two models and a comparison of these would benefit practicing managers to make a decision on the choice of the models.

We use the classical matrix-analytic method (see, e.g., [30]) to formulate the associated stochastic process with matrix-geometric steady-state probability vector. Performance measures of interest are obtained in terms of the steady-state probabilities.

The structure of the present paper is as follows. Models 1 and 2 are analyzed in Sections 3 and 4, respectively. Some selected system performance measures are presented in Sections 3.2 and 4.2. In Section 5, we numerically study the key performance measures of both models in steady-state as well as present a cost analysis to compare the two models. Some concluding remarks are noted in Section 6. The following notation is used in the paper.

- The subscript $i, i = 1, 2$ refers to the model under consideration.
- The symbol $'$ stands for the transpose notation.
- The notations $\otimes$ and $\oplus$ stand for the Kronecker product and Kronecker sum, respectively (see, e.g., [32,33]).

- $\boldsymbol{e'} = (1, 1, \cdots, 1)$, whose dimension should be clear in the context. Where more clarity is needed, the dimension will be mentioned, e.g., $\boldsymbol{e}(m)$ is a column vector of 1s of dimension m.
- $\boldsymbol{e'}_i = (0, 0, \cdots, 1, 0, \cdots, 0)$, where 1 is in the ith position.
- I denotes an identity matrix, whose dimension is dictated by the context.
- $\Delta(E_1, \cdots, E_r)$ denotes a diagonal matrix with diagonal (possibly block) entries given by E_1 through E_r. In the context where this notation is used, it will be clear whether the entities are scalars or vectors or matrices.

3. Opportunistic Model 1

In this first model, we consider the scenario where the customers arrive to a service facility consisting of a single server. The arrivals are modeled using Neuts' versatile Markovian point process, now referred to as batch Markovian arrival process ($BMAP$). While the customers arrive singly to the service facility, their demands for inventory items are in varying sizes. The MAP parameter matrices describing the arrivals of customers are (E_0, E_1) of dimension m. The customers' demands are in batches of varying sizes. Suppose that M denotes the demand size. We assume that M has the following discrete distribution on finite support,

$$P(M = i) = p_i,\ 1 \leq i \leq N, \quad \sum_{i=1}^{N} p_i = 1. \tag{1}$$

In order to determine the arrival rate of the customers to the system, we first define the invariant vector of the generator $E = E_0 + E_1$ that governs the underlying Markov process for the arrivals. Suppose that $\boldsymbol{\theta}$ is the invariant probability vector of E. That is,

$$\boldsymbol{\theta} E = \mathbf{0}, \quad \boldsymbol{\theta}\boldsymbol{e} = 1. \tag{2}$$

Then, the arrival rate of the customers is given by

$$\lambda = \boldsymbol{\theta} E_1 \boldsymbol{e}. \tag{3}$$

Let μ_M to be the mean of the demand size of each customer. That is,

$$\mu_M = \sum_{i=1}^{N} i\, p_i. \tag{4}$$

In this paper, we assume that the demands of the customers are either fully met or partially met or not met at all, depending on, respectively, if the demands in the inventory are fully adequate, partially adequate, or the inventory level is zero. When the inventory level is zero at the time of the arrival of a customer, the arriving customer will be lost.

The size of the inventory is assumed to be finite with a capacity to hold at most K items. As described and motivated earlier, the main purpose of this paper was to incorporate opportunistic replenishment into stochastic models dealing with queues with inventory and positive services. The opportunistic events occur according to a Poisson process with parameter γ. Therefore, when an opportunity occurs for replenishing the inventory, the system has the ability to bring the current inventory level to the maximum level through a given probability structure. Suppose that the inventory level at the time a replenishment opportunity is i. Then, the opportunity will be availed to bring the inventory level to K with probability 1 if $i \leq L$; however, if $i > L$, then with probability $a_{i-L},\ 0 \leq a_{i-L} < 1$, the opportunity will be availed to bring the inventory to K, for $L < i \leq K - 1$.

When a customer enters the system (either with fully or partially met demands), the inventory level will be reduced by that customer's demand size. Note that the only way an arriving customer will be lost is when the inventory level is zero at that moment. However, customers have to wait until their demands are processed with a positive service time, which is assumed to be random. The services follow a phase-type (PH)-distribution with representation $(\boldsymbol{\beta}, S)$ of order n. Note that the mean service rate, say, μ, is given by

$$\mu = \left[\boldsymbol{\beta}(-S)^{-1}\boldsymbol{e}\right]^{-1}. \tag{5}$$

For use in sequel, we define $\boldsymbol{S^0} = -S\boldsymbol{e}$.

Suppose we define the quantities:

- $J_1(t)$ to be the number of customers in the system (including one in service),
- $J_2(t)$ to be the level of the inventory,
- $J_3(t)$ to be the phase of the service (if the server is idle, this will not be defined),
- $J_4(t)$ to be the phase of the arrival process,

at time t. The Markov process $\{(J_1(t), J_2(t), J_3(t), J_4(t)) : t \geq 0\}$ is a quasi-birth-and-death process (QBD) with state space given by

$$\begin{gathered}\Omega_1 = \{(0, j, r) : 0 \leq j \leq K,\ 1 \leq r \leq m\} \\ \bigcup\{((i, j, k, r) : 0 \leq j \leq K,\ 1 \leq k \leq n,\ 1 \leq r \leq m,\ i \geq 1\}.\end{gathered}$$

We now define the levels based on the set of states defined above. Suppose that $\boldsymbol{0} = \{(0, j, r) : 0 \leq j \leq K,\ 1 \leq r \leq m\}$ and $\boldsymbol{i} = \{(i, j, k, r) : 0 \leq j \leq K,\ 1 \leq k \leq n,\ 1 \leq r \leq m\},\ i \geq 1$. This way, we notice that $\boldsymbol{0}$ corresponds to an idle system while $\boldsymbol{i}$ corresponds to the system having i customers with one in service, and the inventory, and the arrival process are in various states. For use in the sequel, we define a few additional notations.

- $\hat{p}_i$, probabilities of demands greater than i, are computed as

$$\hat{p}_i = \sum_{k=i+1}^{N} p_k,\ 1 \leq i \leq N-1. \tag{6}$$

- P, a square matrix of dimension $K+1$ is defined as

$$P = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \hat{p}_1 & p_1 & 0 & \cdots & 0 & 0 & 0 \\ \hat{p}_2 & p_2 & p_1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \hat{p}_{N-1} & p_{N-1} & p_{N-2} & \cdots & 0 & 0 & 0 \\ 0 & p_N & p_{N-1} & \cdots & 0 & 0 & 0 \\ 0 & 0 & p_N & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & p_2 & p_1 & 0 \end{bmatrix}. \tag{7}$$

A quick look at P indicates that the customers' demands are taken into consideration at the time of their arrivals. The first column of P justifies that the structure due to the customers' demands are met partially. It is easy to verify that

$$P\boldsymbol{e} = (0, 1, \ldots, 1)'. \tag{8}$$

The generator, Q_1, of the QBD-process for Opportunistic Model 1 is given by (note that blank spaces correspond to the entries are zero)

$$Q_1 = \begin{pmatrix} B & P \otimes \boldsymbol{\beta} \otimes E_1 & & & & \\ I \otimes \boldsymbol{S^0} \otimes I & A_1 & A_0 & & & \\ & A_2 & A_1 & A_0 & & \\ & & A_2 & A_1 & A_0 & \\ & & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{9}$$

For a better display of the matrices, we take

$$\begin{array}{lll} F & = & E - \gamma I, \\ F_0 & = & E_0 - \gamma I, \\ F_r & = & E_0 - a_r \gamma I,\ 1 \le r \le K - L - 1, \\ S_E & = & (S \oplus E) - \gamma I, \\ S_{E_0} & = & (S \oplus E_0) - \gamma I, \\ S_{E_r} & = & (S \oplus E_0) - a_r \gamma I,\ 1 \le r \le K - L - 1. \end{array} \tag{10}$$

The matrices appearing in Q_1 are as follows.

$$B = \begin{array}{c} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{2} \\ \vdots \\ \boldsymbol{L} \\ \boldsymbol{L+1} \\ \vdots \\ \boldsymbol{K-1} \\ \boldsymbol{K} \end{array} \begin{array}{c} \begin{array}{ccccccccc} \mathbf{0} & \mathbf{1} & \mathbf{2} & \cdots & \boldsymbol{L} & \boldsymbol{L+1} & \cdots & \boldsymbol{K-1} & \boldsymbol{K} \end{array} \\ \begin{pmatrix} F & & & & & & & & \gamma I \\ & F_0 & & & & & & & \gamma I \\ & & F_0 & & & & & & \gamma I \\ & & & \ddots & & & & & \vdots \\ & & & & F_0 & & & & \gamma I \\ & & & & & F_1 & & & a_1 \gamma I \\ & & & & & & \ddots & & \vdots \\ & & & & & & & F_{K-L-1} & a_{K-L-1} \gamma I \\ & & & & & & & & E_0 \end{pmatrix} \end{array}, \tag{11}$$

$$A_1 = \begin{array}{c} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{2} \\ \vdots \\ \boldsymbol{L} \\ \boldsymbol{L+1} \\ \vdots \\ \boldsymbol{K-1} \\ \boldsymbol{K} \end{array} \begin{array}{c} \begin{array}{ccccccccc} \mathbf{0} & \mathbf{1} & \mathbf{2} & \cdots & \boldsymbol{L} & \boldsymbol{L+1} & \cdots & \boldsymbol{K-1} & \boldsymbol{K} \end{array} \\ \begin{pmatrix} S_E & & & & & & & & \gamma I \\ & S_{E_0} & & & & & & & \gamma I \\ & & S_{E_0} & & & & & & \gamma I \\ & & & \ddots & & & & & \vdots \\ & & & & S_{E_0} & & & & \gamma I \\ & & & & & S_{E_1} & & & a_1 \gamma I \\ & & & & & & \ddots & & \vdots \\ & & & & & & & S_{E_{K-L-1}} & a_{K-L-1} \gamma I \\ & & & & & & & & S_{E_0} + \gamma I \end{pmatrix} \end{array}, \tag{12}$$

and

$$A_0 = P \otimes I \otimes E_1, \ \ A_2 = I \otimes \boldsymbol{S^0} \boldsymbol{\beta} \otimes I. \tag{13}$$

Suppose that $\boldsymbol{\eta} = (\boldsymbol{\eta}_0, \cdots, \boldsymbol{\eta}_K)$ denotes the invariant probability vector of $A = A_0 + A_1 + A_2$, that is, $\boldsymbol{\eta} A = 0$ and $\boldsymbol{\eta e} = 1$. The following lemma, which is intuitively clear, is very useful in numerical computation as well as for probabilistic interpretation.

Lemma 1. *The vector $\boldsymbol{\eta}$ is such that*

$$\sum_{i=0}^{K} \boldsymbol{\eta}_i = (\mu \boldsymbol{\beta}(-S^{-1}) \otimes \boldsymbol{\theta}). \tag{14}$$

Proof. Setting $A = A_0 + A_1 + A_2, \tilde{S} = S + \boldsymbol{S^0}\boldsymbol{\beta}$, and noting that $A(\boldsymbol{e} \otimes I \otimes I) = [\boldsymbol{e} \otimes (\tilde{S} \oplus E)]$, we see that

$$\boldsymbol{\eta} A = \mathbf{0} \;\Rightarrow\; \boldsymbol{\eta} A(\boldsymbol{e} \otimes I \otimes I) = \mathbf{0} \;\Rightarrow\; \boldsymbol{\eta}[\boldsymbol{e} \otimes (\tilde{S} \oplus E)] = \mathbf{0}, \tag{15}$$

which implies that

$$\boldsymbol{\eta}(\boldsymbol{e} \otimes I)[\tilde{S} \oplus E] = \mathbf{0}. \tag{16}$$

On noting that the the invariant vectors of $\tilde{S}$ and E, are, respectively, given by $\mu\boldsymbol{\beta}(-S^{-1})$ and $\boldsymbol{\theta}$, the stated result follows from Equation (16).

The computation of $\boldsymbol{\eta}$ is done by exploiting the structure of the matrices appearing in A. From the definition of $\boldsymbol{\eta}$ and the equations associated with the invariant vector, we get

$$\begin{array}{lcl} \boldsymbol{\eta}_0 & = & \sum_{j=1}^{N} \boldsymbol{\eta}_j p_{j,0}(I \otimes E_1)[\gamma I - (\tilde{S} \oplus E)]^{-1}, \\ \boldsymbol{\eta}_k & = & \sum_{j=k+1}^{N+k} \boldsymbol{\eta}_k p_{j,k}(I \otimes E_1)[\gamma I - (\tilde{S} \oplus E_0)]^{-1},\ 1 \le k \le L, \\ \boldsymbol{\eta}_k & = & \sum_{j=k+1}^{N+k} \boldsymbol{\eta}_k p_{j,k}(I \otimes E_1)[a_{k-L}\,\gamma I - (\tilde{S} \oplus E_0)]^{-1},\ L+1 \le k \le K-1, \\ \boldsymbol{\eta}_K & = & \gamma\,[\sum_{j=0}^{L} \boldsymbol{\eta}_k + \sum_{j=L+1}^{K-1} a_{j-L}\,\boldsymbol{\eta}_k](-(\tilde{S} \oplus E_0))^{-1}. \end{array} \tag{17}$$

The following lemma, which again is intuitively clear, gives necessary and sufficient condition for the stability of the model under study. □

Lemma 2. *The queuing-inventory model with opportunistic replenishment with the generator given in (9) is stable if and only if*

$$\lambda - \boldsymbol{\eta}_0(\boldsymbol{e} \otimes E_1\boldsymbol{e}) < \mu. \tag{18}$$

Proof. From the celebrated Neuts' ergodicity criterion (see, e.g., [30]) for stability of QBD-process: $\pi A_0\boldsymbol{e} < \pi A_2\boldsymbol{e}$, the stated result follows immediately after some elementary matrix manipulations.

Letting $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \cdots)$, with the vector $\boldsymbol{x}_0$ of dimension $(K+1)m$ and each of the vectors $\boldsymbol{x}_i,\ i \ge 1$, of dimension $(K+1)mn$, be the steady-state probability vector of $Q^{(1)}$ process under consideration, we notice the following:

$$\boldsymbol{x}Q_1 = \mathbf{0},\ \ \boldsymbol{x}\boldsymbol{e} = 1. \tag{19}$$

The following theorem is a direct consequence of Neuts' result on QBD-process (see, e.g., [30]) and we state it here for the sake of completeness. □

Theorem 1. *Under the stability condition given in (18), the steady-state probability vector, $\boldsymbol{x}$, of Q_1 is of matrix-geometric type:*

$$\boldsymbol{x}_i = \boldsymbol{x}_1 R^{i-1},\ i \ge 1, \tag{20}$$

where the (rate) matrix R is the minimal non-negative solution to:

$$R^2 A_2 + RA_1 + A_0 = 0. \tag{21}$$

and the vectors $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ are obtained by solving

$$\begin{array}{rcl} \boldsymbol{x}_0 B + \boldsymbol{x}_1(I \otimes \boldsymbol{S^0} \otimes I) & = & \mathbf{0}, \\ \boldsymbol{x}_0(P \otimes \boldsymbol{\beta} \otimes E_1) + \boldsymbol{x}_1(A_1 + RA_2) & = & \mathbf{0}, \end{array} \tag{22}$$

subject to:

$$\boldsymbol{x}_0\boldsymbol{e} + \boldsymbol{x}_1(I - R)^{-1}\boldsymbol{e} = 1. \tag{23}$$

3.1. Computation of R

Without any additional special structure to the rate matrix, R, it is pertinent that we compute it numerically. Of course, any sparsity in the coefficient matrices seen in the matrix-quadratic equation given in (21) should be fully exploited (see, e.g., [20,21,30,34]), especially, when m, n, L, and K are sufficiently large. In addition, the fact that $R\, A_2\, \boldsymbol{e} A_0\, \boldsymbol{e}$, will help to identify if there are any zero rows in R further simplifying the computational steps. For example, let $V = R^2$ and partition R into matrices of dimension mn as follows.

$$R = \begin{pmatrix} R_{0,0} & R_{0,1} & \cdots & R_{0,K} \\ R_{0,0} & R_{0,1} & \cdots & R_{0,K} \\ \vdots & \vdots & \cdots & \vdots \\ R_{K,0} & R_{K,1} & \cdots & R_{K,K} \end{pmatrix}. \tag{24}$$

Noting that for the model under study $R_{0,j}, 0 \leq j \leq K = 0$, we can exploit that observation. Now, the equation in (21) with the help of the partition of R can be rewritten for numerical implementation as follows, noting that, for $1 \leq i \leq K$,

$$\begin{aligned} V_{i,0}(\boldsymbol{S^0}\boldsymbol{\beta} \otimes I) + R_{i,0}(S \oplus E - \gamma I) + p_{i,0}(I \otimes E_1) &= 0, \\ V_{i,j}(\boldsymbol{S^0}\boldsymbol{\beta} \otimes I) + R_{i,j}(S \oplus E_0 - \gamma I) + p_{i,j}(I \otimes E_1) &= 0, \quad 1 \leq j \leq L, \\ V_{i,j}(\boldsymbol{S^0}\boldsymbol{\beta} \otimes I) + R_{i,j}(S \oplus E_0) + p_{i,j}(I \otimes E_1) &= 0, \quad L+1 \leq j \leq K-1, \\ V_{i,K}(\boldsymbol{S^0}\boldsymbol{\beta} \otimes I) + R_{i,K}(S \oplus E_0) + \gamma R_{i,L} + p_{i,K}(I \otimes E_1) &= 0. \end{aligned} \tag{25}$$

If m and n are very large, one can further exploit the coefficient matrices seen in (25). For example, the Kronecker products and sums appearing in the Equation (25) can be exploited so as to deal only with matrices of dimension m. Suppose that

$$\zeta_{i,j} = \lim_{t \to \infty} \mathrm{P}(J_2(t) = i, J_4(t) = j) =, \; 0 \leq i \leq K, \; 1 \leq j \leq m. \tag{26}$$

Recall that $J_2(t)$, and $J_4(t)$ are, respectively, track of the number in the inventory and the phase of the arrival process. Let $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_0, \boldsymbol{\zeta}_1, \cdots, \boldsymbol{\zeta}_K)$ be such that $\boldsymbol{\zeta}_i$, $0 \leq i \leq K$, of dimension m gives the (marginal) probability vector of seeing i in the inventory with the phase of the arrival process being in various phases in steady-state.

Lemma 3. *The (marginal) probability vector $\boldsymbol{\zeta}$ is independent of the service time distribution.*

Proof. Observe that the steady-state equations given in (36) when rewritten yield

$$\begin{aligned} \boldsymbol{x}_0 B + \boldsymbol{x}_1(I \otimes \boldsymbol{S^0} \otimes I) &= \boldsymbol{0}, \\ \boldsymbol{x}_0(P \otimes \boldsymbol{\beta} \otimes E_1) + \boldsymbol{x}_1 A_1 + \boldsymbol{x}_2 A_2 &= \boldsymbol{0}, \\ \boldsymbol{x}_{i-1} A_0 + \boldsymbol{x}_i A_1 + \boldsymbol{x}_{i+1} A_2 &= \boldsymbol{0}, \; i \geq 2. \end{aligned} \tag{27}$$

Suppose we define $H = B + P \otimes E_1$, a square matrix of dimension $(K+1)mn$, and partition it into blocks of dimension $mn \times mn$ as

$$H = \begin{pmatrix} H_{0,0} & H_{0,1} & \cdots & H_{0,K} \\ H_{1,0} & H_{1,1} & \cdots & H_{1,K} \\ \vdots & \vdots & \vdots & \vdots \\ H_{K,0} & H_{K,1} & \cdots & H_{K,K} \end{pmatrix}. \tag{28}$$

From the definitions of B (see Equation (11)) and P (see Equation (7)), it is easy to identify the (block) elements of H.

It is easy to verify that $A(I \otimes e \otimes I)$ is given by

$$\begin{pmatrix} e \otimes H_{0,0} & e \otimes H_{0,1} & \cdots & e \otimes H_{0,K} \\ e \otimes H_{1,0} & e \otimes H_{1,1} & \cdots & e \otimes H_{1,K} \\ \vdots & \vdots & \vdots & \vdots \\ e \otimes H_{K,0} & e \otimes H_{K,1} & \cdots & e \otimes H_{K,K} \end{pmatrix}. \tag{29}$$

By post-multiplying the second and third equations in (27) by $I \otimes e \otimes I$ and adding the resulting equations with the first equation in (27), we get

$$x_0 H + \sum_{i=1}^{\infty} x_i A(I \otimes e \otimes I) = \mathbf{0}, \tag{30}$$

which, on observing that

$$\zeta = x_0 + \sum_{i=1}^{\infty} x_i (I \otimes e \otimes I), \tag{31}$$

implies that

$$\zeta H = \mathbf{0}. \tag{32}$$

The above equation yields the stated result as the matrix H is independent of the service time distribution. □

Remark 1. *Lemma 3 is intuitively clear. This is due to the fact that the inventory level is decreased at arrival times and the service time plays no role.*

3.2. Selected System Measures in Steady-State

For qualitative evaluation of the models presented in this paper, we look at the system performance measures in the following table. For Opportunistic Model 1, the measures along with their formulas are as follows.

1. Server idle probability: $\nu_1 = \mathbf{x_0} e$.
2. Probability of idle server with positive inventory: $\nu_{1I} = \mathbf{x_0} e - \mathbf{x_{0,0}} e$.
3. Percent of server idle time with positive inventory: $\nu_1^* = \nu_{1I}/\nu_1$
4. Mean number of customers in the system: $\mu_1 = \mathbf{x_1}(I - R)^{-2} e$.
5. Variance of the number of customers in the system:

$$\sigma_1^2 = 2\mathbf{x_1}(I - R)^{-3} e - \mu_1(1 + \mu_1).$$

6. Probability of customer loss (at arrivals): $\vartheta_1 = \frac{1}{\lambda}\left[\sum_{i=0}^{\infty} \mathbf{x_{i,0}}(e \otimes E_1 e)\right]$.
7. Mean inventory level: $\hat{\mu}_1 = \sum_{j=1}^{K} j\left[\mathbf{x_{0,j}} e + \sum_{i=1}^{\infty} \mathbf{x_{i,j}} e\right]$.
8. Variance of the inventory level: $\hat{\sigma}_1^2 = \sum_{j=1}^{K} j^2\left[\mathbf{x_{0,j}} e + \sum_{i=1}^{\infty} \mathbf{x_{i,j}} e\right] - \hat{\mu}_1^2$.
9. Mean cycle time of replenishment:

$$\kappa_1 = \left(\gamma\left[\sum_{j=0}^{L} \mathbf{x_{0,j}} e + \sum_{j=L+1}^{K-1} a_{j-L} \sum_{i=1}^{\infty} \mathbf{x_{i,j}} e\right]\right)^{-1}$$

10. Mean replenishment quantity:

$$\Gamma_1 = \gamma\left[\sum_{j=0}^{L} (K-j)\left(\mathbf{x_{0,j}} e + \sum_{i=1}^{\infty} \mathbf{x_{i,j}} e\right) + \sum_{j=L+1}^{K-1} (K-j)\, a_{j-L}\left(\mathbf{x_{0,j}} e + \sum_{i=1}^{\infty} \mathbf{x_{i,j}}\, e\right)\right].$$

11. Probability of procuring an order when a replenishment opportunity arises:

$$\xi_1 = \sum_{j=0}^{L} \Big(\boldsymbol{x_{0,j} e} + \sum_{i=1}^{\infty} \boldsymbol{x_{i,j} e} \Big) + \sum_{j=L+1}^{K-1} a_{j-L} \Big(\boldsymbol{x_{0,j} e} + \sum_{i=1}^{\infty} \boldsymbol{x_{i,j}\, e} \Big).$$

4. Opportunistic Model 2

In the Opportunistic Model 1, we assumed that the customers are lost at arrivals when the inventory level is zero. However, in Opportunistic Model 2, we will admit customers into the system even if the inventory is zero as long as the server is busy serving. If the server is idle with zero inventory, then the arriving customers will be lost. In this model, we assume that the inventory level is reduced by the amount the customers' demand is either fully or partially met just before the service starts (or equivalently just after a service completion). Thus, there is a possibility that the customers may be lost at service completion epochs when the inventory level is zero. All the customers waiting at the time of a service completion which results in zero inventory will be lost. This is due to the fact that a service requires at least one inventory. Arrivals, services, and the probabilistic structure for availing the opportunistic events are all the same as in the Opportunistic Model 1.

The Opportunistic Model 2 is of the $GI/M/1$-type [30]. Suppose that we define

- $\hat{J}_1(t)$ to be the number of customers in the system,
- $\hat{J}_2(t)$ to be the level of the inventory,
- $\hat{J}_3(t)$ to be the phase of the service (if the server is idle, this will not be defined),
- $\hat{J}_4(t)$ to be the phase of the arrival process,

at time t. The stochastic process $\{(\hat{J}_1(t), \hat{J}_2(t), \hat{J}_3(t), \hat{J}_4(t)) : t \geq 0\}$ is a Markov process with the state space given by

$$\Omega_2 = \{(0, j, r) : 0 \leq j \leq K,\ 1 \leq r \leq m\}$$

$$\bigcup \{((i, j, k, r) : 0 \leq j \leq K,\ 1 \leq j \leq n,\ 1 \leq r \leq m,\ i \geq 1\}.$$

We define the set of states similar to Opportunistic Model 1. The generator, Q_2, of the system governing Model 2 is of the form

$$Q_2 = \begin{pmatrix} B & P \otimes \boldsymbol{\beta} \otimes E_1 & & & & \\ I \otimes \boldsymbol{S^0} \otimes I & G_1 & G_0 & & & \\ C & G_2 & G_1 & G_0 & & \\ C & & G_2 & G_1 & G_0 & \\ \vdots & & & \ddots & \ddots & \ddots \end{pmatrix}.$$

The matrix B is as given (11) and the matrices C, G_0, G_1, and G_2 appearing in Q_2 are as follows.

$$\begin{aligned} C &= \boldsymbol{e}_1(K+1) \otimes \boldsymbol{e}_1'(K+1) \otimes \boldsymbol{S^0} \otimes I, \\ G_0 &= I \otimes D_1, \quad G_2 = P \otimes \boldsymbol{S^0} \boldsymbol{\beta} \otimes I, \end{aligned} \tag{33}$$

and

$$G_1 = \begin{array}{c} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{2} \\ \vdots \\ \boldsymbol{L} \\ \boldsymbol{L+1} \\ \vdots \\ \boldsymbol{K-1} \\ \boldsymbol{K} \end{array} \begin{array}{c} \begin{array}{ccccccccc} \mathbf{0} & \mathbf{1} & \mathbf{2} & \cdots & \boldsymbol{L} & \boldsymbol{L+1} & \cdots & \boldsymbol{K-1} & \boldsymbol{K} \end{array} \\ \begin{pmatrix} S_{E_0} & & & & & & & & \gamma I \\ & S_{E_0} & & & & & & & \gamma I \\ & & S_{E_0} & & & & & & \gamma I \\ & & & \ddots & & & & & \vdots \\ & & & & S_{E_0} & & & & \gamma I \\ & & & & & S_{E_1} & & & a_1 \gamma I \\ & & & & & & \ddots & & \vdots \\ & & & & & & & S_{E_{K-L-1}} & a_{K-L-1}\gamma I \\ & & & & & & & & S_{E_0} + \gamma I \end{pmatrix} \end{array}, \quad (34)$$

where S_{E_r}, $0 \le r \le K - L - 1$ are as defined in (10). Note that the level **0** can be reached from any other level and hence this queuing model can be thought as a catastrophic model, and hence is always stable. In addition, observe that when $\gamma \to \infty$ the model becomes unstable when $\lambda \ge \mu$.

Suppose that $\boldsymbol{y}$, partitioned as $\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, \cdots)$, satisfies

$$\boldsymbol{y} Q_2 = \mathbf{0}, \;\; \boldsymbol{y}\boldsymbol{e} = 1. \quad (35)$$

The following theorem is a direct consequence of Neuts' result (see, e.g., [30]) and for the sake of completeness, we will register it here.

Theorem 2. *Assuming that γ is finite, the steady-state probability vector, $\boldsymbol{y}$, of Q_2 is of matrix-geometric type and is given by:*

$$\boldsymbol{y}_i = \boldsymbol{y}_1 \hat{R}^{i-1}, \; i \ge 1, \quad (36)$$

where the (rate) matrix $\hat{R}$ is the minimal non-negative solution to:

$$\hat{R}^2 G_2 + \hat{R} G_1 + G_0 = 0. \quad (37)$$

and the vectors $\boldsymbol{y}_0$ and $\boldsymbol{y}_1$ are obtained by solving

$$\begin{aligned} \boldsymbol{y}_0 B + \boldsymbol{y}_1[(I \otimes \boldsymbol{S^0} \otimes I) + \hat{R}(I - \hat{R})^{-1} C] &= \mathbf{0}, \\ \boldsymbol{y}_0 (P \otimes \boldsymbol{\beta} \otimes E_1) + \boldsymbol{y}_1 [G_1 + \hat{R} G_2] &= \mathbf{0}, \end{aligned} \quad (38)$$

subject to

$$\boldsymbol{y}_0\, \boldsymbol{e} + \boldsymbol{y}_1 (I - \hat{R})^{-1} \boldsymbol{e} = 1.$$

The rate matrix $\hat{R}$ can be efficiently computed by exploiting the structure of the coefficient matrices and the steps are similar to the computation of R seen in the Opportunistic Model 1 and hence will be omitted. Again, the exploitation of the structure of the coefficient matrices will result in dealing with matrices of smaller dimensions.

4.1. Computation of $\boldsymbol{y}_0$ and $\boldsymbol{y}_1$

Due to the special structure of the coefficient matrices appearing in the solution of the vectors, $\boldsymbol{y}_0$ and $\boldsymbol{y}_1$, it is worth to exploit them as follows. First, we partition these vectors as follows.

$$\boldsymbol{y}_0 = (\boldsymbol{y}_{0,0}, \cdots, \boldsymbol{y}_{0,K}), \;\; \boldsymbol{y}_1 = (\boldsymbol{y}_{1,0}, \cdots, \boldsymbol{y}_{1,K}). \quad (39)$$

Note that the vectors, $\boldsymbol{y}_{0,j}$, $0 \leq j \leq K$, have dimension m, whereas the vectors, $\boldsymbol{y}_{1,j}$, $0 \leq j \leq K$, have dimension mn.

$$\begin{aligned}
\boldsymbol{y}_{0,0} &= \boldsymbol{y}_1 (I-\hat{R})^{-1} \begin{pmatrix} \boldsymbol{S^0} \otimes I \\ 0 \end{pmatrix} (\gamma I - E)^{-1}, \\
\boldsymbol{y}_{0,j} &= \textstyle\sum_{k=1}^{n} \boldsymbol{y}_{1,j,k} S_k^0 (\gamma I - E_0)^{-1}, \ 1 \leq j \leq L, \\
\boldsymbol{y}_{0,j} &= \textstyle\sum_{k=1}^{n} \boldsymbol{y}_{1,j,k} S_k^0 (a_{j-L} \gamma I - E_0)^{-1}, \ L+1 \leq j \leq K-1, \\
\boldsymbol{y}_{0,K} &= \left[\gamma \textstyle\sum_{k=0}^{L} \boldsymbol{y}_{0,k} + \gamma \sum_{k=L+1}^{K-1} a_{k-L} \boldsymbol{y}_{0,k} + \sum_{k=1}^{n} \boldsymbol{y}_{1,j,k} S_k^0 S_k^0 \right] (-E_0)^{-1}.
\end{aligned} \tag{40}$$

$$\begin{aligned}
\boldsymbol{y}_{1,j} &= \left[\textstyle\sum_{k=0}^{K} p_{k,j} (\boldsymbol{\beta} \otimes E_1) + \sum_{k=0}^{K} \boldsymbol{y}_{1,k} D_{k,j}\right] [\gamma I - (S \oplus E_0)]^{-1}, 0 \leq j \leq L, \\
\boldsymbol{y}_{1,j} &= \left[\textstyle\sum_{k=0}^{K} p_{k,j} (\boldsymbol{\beta} \otimes E_1) + \sum_{k=0}^{K} \boldsymbol{y}_{1,k} D_{k,j}\right] [a_{j-L}\, \gamma I - (S \oplus E_0)]^{-1}, L+1 \leq j \leq K-1, \\
\boldsymbol{y}_{1,K} &= \left[\gamma \textstyle\sum_{k=0}^{L} \boldsymbol{y}_{1,k} + \gamma \sum_{k=L+1}^{K-1} a_{k-L} \boldsymbol{y}_{1,k} + \sum_{k=0}^{K} \boldsymbol{y}_{1,j,k} D_{k,j}\right] [-(S \oplus E_0)]^{-1},
\end{aligned} \tag{41}$$

and the normalizing constant is given by

$$\sum_{k=0}^{K} \boldsymbol{y}_0\, \boldsymbol{e} + \sum_{k=0}^{K} \boldsymbol{y}_1\, \boldsymbol{b} = 1, \tag{42}$$

where the block entries, $D_{j,k}$, $0 \leq j,k \leq K$, of the matrix D and the vector $\boldsymbol{b}$ are defined as

$$\boldsymbol{b} = (I-\hat{R})^{-1} \boldsymbol{e}, \ \ \hat{R} G_2 = D. \tag{43}$$

The vector $\boldsymbol{b} = (\boldsymbol{b}_0, \cdots, \boldsymbol{b}_K)$ is obtained as follows. Suppose that $(I-\hat{R})\, \boldsymbol{b} = \boldsymbol{e}((K+1)mn)$, then we have

$$\boldsymbol{b}_i = (I - \hat{R}_{i,i})^{-1} \left[\boldsymbol{e}(mn) + \sum_{j=0, j\neq i}^{K} \hat{R}_{i,j} \boldsymbol{b}\right]. \tag{44}$$

The block entries of D are computed as follows

$$D_{j,k} = \sum_{k=0}^{K} p_{k,j}\, \hat{R}_{i,k}\, (\boldsymbol{S^0} \boldsymbol{\beta} \otimes I), \ 0 \leq j,k \leq K. \tag{45}$$

Note that the quantity

$$(I-\hat{R})^{-1} \begin{pmatrix} \boldsymbol{S^0} \otimes I \\ 0 \end{pmatrix} = (I-\hat{R})^{-1} \begin{pmatrix} \boldsymbol{S^0} \otimes \boldsymbol{e}_1(m) & \boldsymbol{S^0} \otimes \boldsymbol{e}_2(m) & \cdots & \boldsymbol{S^0} \otimes \boldsymbol{e}_m(m) \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \tag{46}$$

can be obtained by solving for $\boldsymbol{d}_j$, $1 \leq j \leq m$, which satisfies

$$(I-\hat{R})\, \boldsymbol{d}_j = \begin{pmatrix} \boldsymbol{S^0} \otimes \boldsymbol{e}_j(m) \\ 0 \end{pmatrix}, \tag{47}$$

and one can use the same type of equations seen in (44) by replacing $\boldsymbol{e}(mn)$ with

$$\begin{pmatrix} \boldsymbol{S^0} \otimes \boldsymbol{e}_j(m) \\ 0 \end{pmatrix}. \tag{48}$$

4.2. Selected System Measures in Steady-State

For qualitative evaluation of the models presented in this paper, we look at the system performance measures in the following table. For the Opportunistic Model 1, the measures along with their formulas are as follows.

1. Server idle probability: $\nu_2 = \boldsymbol{y_0}\boldsymbol{e}$.
2. Probability of idle server with positive inventory: $\nu_{2I} = \boldsymbol{y_0}\boldsymbol{e} - \boldsymbol{y_{0,0}}\boldsymbol{e}$.
3. Percent of server idle time with positive inventory: $\nu_2^* = \nu_{2I}/\nu_2$.
4. Mean number of customers in the system: $\mu_2 = \boldsymbol{y_1}(I - \hat{R})^{-2}\boldsymbol{e}$.
5. Variance of the number of customers in the system:
$$\sigma_2^2 = 2\boldsymbol{y_1}(I - \hat{R})^{-3}\boldsymbol{e} - \mu_2(1 + \mu_2).$$
6. Probability of customer loss at arrivals:
$$\vartheta_2 = \frac{1}{\lambda}\Big[\boldsymbol{y_{0,0}}(\boldsymbol{e} \otimes E_1\boldsymbol{e}) + \sum_{i=1}^{\infty}(i-1)\boldsymbol{y_{i,0}}(\boldsymbol{S^0} \otimes \boldsymbol{e})\Big].$$
7. Mean inventory level: $\hat{\mu}_2 = \sum_{j=1}^{K} j\Big[\boldsymbol{y_{0,j}}\boldsymbol{e} + \sum_{i=1}^{\infty}\boldsymbol{y_{i,j}}\boldsymbol{e}\Big]$.
8. Variance of the inventory level: $\hat{\sigma}_2^2 = \sum_{j=1}^{K} j^2\Big[\boldsymbol{y_{0,j}}\boldsymbol{e} + \sum_{i=1}^{\infty}\boldsymbol{y_{i,j}}\boldsymbol{e}\Big] - \hat{\mu}_2^2$.
9. Mean cycle time of replenishment:
$$\kappa_2 = \Big(\gamma\Big[\sum_{j=0}^{L}\boldsymbol{y_{0,j}}\boldsymbol{e} + \sum_{j=L+1}^{K-1} a_{j-L}\sum_{i=1}^{\infty}\boldsymbol{y_{i,j}}\boldsymbol{e}\Big]\Big)^{-1}.$$
10. Mean replenishment quantity:
$$\Gamma_2 = \gamma\Big[\sum_{j=0}^{L}(K-j)\Big(\boldsymbol{y_{0,j}}\boldsymbol{e} + \sum_{i=1}^{\infty}\boldsymbol{y_{i,j}}\boldsymbol{e}\Big) + \sum_{j=L+1}^{K-1}(K-j)\,a_{j-L}\Big(\boldsymbol{y_{0,j}}\boldsymbol{e} + \sum_{i=1}^{\infty}\boldsymbol{y_{i,j}}\,\boldsymbol{e}\Big)\Big].$$
11. Probability of procuring an order when a replenishment opportunity arises:
$$\xi_2 = \sum_{j=0}^{L}\Big(\boldsymbol{y_{0,j}}\boldsymbol{e} + \sum_{i=1}^{\infty}\boldsymbol{y_{i,j}}\boldsymbol{e}\Big) + \sum_{j=L+1}^{K-1} a_{j-L}\Big(\boldsymbol{y_{0,j}}\boldsymbol{e} + \sum_{i=1}^{\infty}\boldsymbol{y_{i,j}}\,\boldsymbol{e}\Big).$$

Note that the loss probability, ϑ_2, has two components, the first is for the loss at arrivals (ϑ_{2a}) and the second one is for the loss at a service completion due to zero inventory (ϑ_{2d}), which makes the server become idle and hence it cannot serve any customer.

5. Illustrative Numerical Examples

First, we describe the experimental setup. We start with considering the dependence of relative difference of performance measures across various batch size distributions.

We define the following arrival processes (along with their parameters including the standard deviations):

ERA Erlang distributed inter-arrival times with density $\tilde{\lambda}^k x^{k-1} e^{-\tilde{\lambda}x}/(k-1)!$, $\tilde{\lambda} = 4$, and $k = 4$ (standard deviation ≈ 0.5774).

HEA Hyperexponential inter-arrival times with density $\sum_{i=1}^{k} p_i\tilde{\lambda}_i e^{-\tilde{\lambda}_i x}$, $k = 4$, $\tilde{\lambda}_i = (63.1)\cdot 10^{-i}, i = 1,2,3$, and $p = (0.6, 0.25, 0.10, 0.05)$ (standard deviation ≈ 4.9629).

PCA Markov arrival process (E_0, E_1) with positive correlation ≈ 0.4637 (standard deviation ≈ 1.3153), where

$$E_0 = \begin{pmatrix} -1.05 & 1.05 & 0 \\ 0 & -1.05 & 0 \\ 0 & 0 & -10.5 \end{pmatrix}, \quad E_1 = \begin{pmatrix} 0 & 0 & 0 \\ 1.035 & 0 & 0.0105 \\ 0.105 & 0 & 10.395 \end{pmatrix}.$$

The parameters of distributions are normalized so as to obtain a unit arrival rate. The graphs of the probability density function of the three inter-arrival times are plotted in Figure 1 and the joint probability function of the *MAP* with positive correlation is plotted in Figure 2.

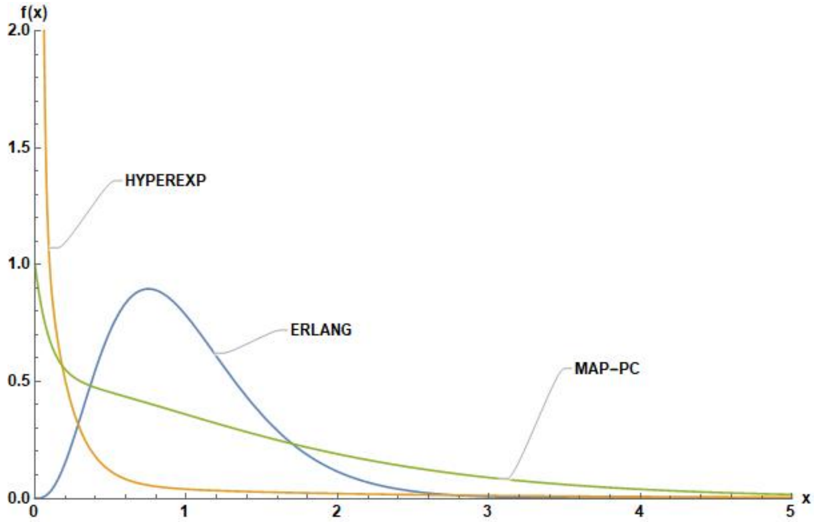

Figure 1. Probability density functions of the inter-arrival times.

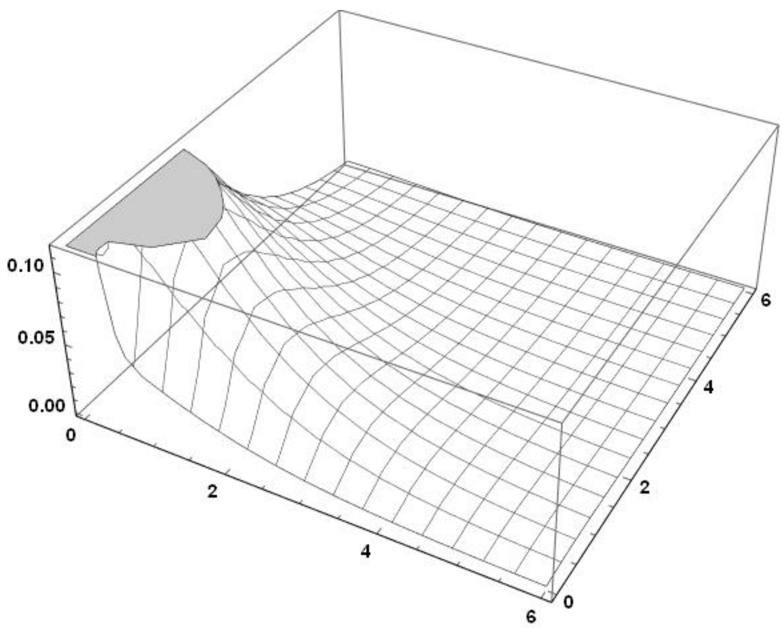

Figure 2. Joint Probability Density Function of $MAP - PC$.

We also use the following service time distributions:

ERS Erlang distributed service times having density $\tilde{\mu}^k x^{k-1} e^{-\tilde{\mu} x}/(k-1)!$ with $k = 3$.

EXS Exponentially distributed service times of rate μ.

HES Hyperexponentially distributed service times with density $\sum_{i=1}^{k} p_i \tilde{\mu}_i e^{-\tilde{\mu}_i x}$, $k = 3$, $p = (0.7, 0.25, 0.05)$ and $\tilde{\mu}_i = (8.2\,\mu) \cdot 10^{-i}, i = 1, \ldots, 3$.

The service rate μ will be normalized so as to arrive at a given service rate. Obviously, the above three PH-distributions are qualitatively different covering a wide range applicable in practice. The plot of these three distributions is given in Figure 3.

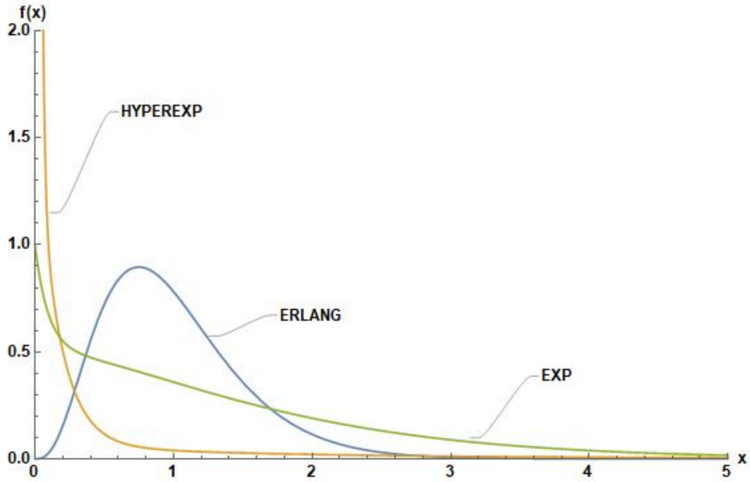

Figure 3. Probability density function of the service times.

For our illustrative examples, here we take the batch size distribution to be uniform in $\{1, \ldots, N\}$ and the batch size maximum to be N, which is set at $N = 7$. It should be pointed out that we did consider other batch size distributions like constant, truncated Poisson, and truncated geometric. For the range of parameters considered, we observed that numerical results with other batch size distribution are similar to those with uniform distribution.

When a replenishment opportunity occurs, an order is placed with probability 1 when the inventory level is at most L, and with probability a_i, when the inventory level is $i, L+1 < i \leq K-1$. Order quantity is always determined to bring the inventory level to K. We considered three types of probability functions for a_i, namely, constant, linearly decreasing, and non-linearly decreasing for $i = L+1 \leq i \leq K-1$. We observed that the effect of the type of probability function used for a_i on the system performance is minimal. In many cases, the corresponding results for the three probability functions are equal within the fourth decimal.

Queuing-inventory systems consist of two interacting subsystems, namely the inventory subsystem (ISS) and the customer subsystem (CSS).

- In ISS, inventory is consumed and replenished as needed and is characterized by the parameters K, L, γ and the distribution of the time between two opportunities for replenishment. The arrival process to CSS impacts ISS through the demand for items. $\hat{\mu}_1[\hat{\mu}_2]$, $\hat{\sigma}_1[\hat{\sigma}_2]$, $\Gamma_1[\Gamma_2]$, $\xi_1[\xi_2]$, $\kappa_1[\kappa_2]$ represent the measures of performance for ISS for Model 1 [Model 2].
- In CSS, customers arrive, receive service (plus items from inventory), and depart and are characterized by the arrival and service processes. Some arrivals are lost due to lack of inventory at the time of arrival. In Model 2, customers may also be lost at a service completion epoch due to lack of inventory at that moment. Customer loss is impacted by the availability of inventory in ISS. $\mu_1[\mu_2]$, $\sigma_1[\sigma_2]$, $\nu_1[\nu_2]$, $\nu_1^*[\nu_2^*]$, $\vartheta_1[\vartheta_{2a}, \vartheta_{2d}]$ represent measures of performance for CSS for Model 1 [Model 2].

Inventory levels control the interaction between the two subsystems. If the inventory levels are high [low], customer loss and the interaction between ISS and CSS will be low [high]. In the limiting case when infinite inventory is maintained, interaction between the two subsystems decreases to zero and Model 2 converges to Model 1. In this paper, we consider $\vartheta_1[\vartheta_2]$ and $\mu_1[\mu_2]$ as good measures for describing the extent of interaction between ISS and CSS for the two models.

Tables 1 and 2 respectively summarize the results for CSS and ISS for a broad range of parameter values. These tables represent a subset of the numerical results used in the following qualitative summary of the effect of various parameters on the performance of two subsystems for the two models. Table 3 compares the system measures for the two models. In these tables, the abbreviations Arr. and Ser. are used to identify the arrival process and service time distribution.

Table 1. Customer subsystem (μ = 1.1).

					Model 1					**Model 2**					
Arr.	**Ser.**	γ	K	L	μ_1	σ_1	ν_1	ν_1^*	ϑ_1	μ_2	σ_2	ν_2	ν_2^*	ϑ_{2a}	ϑ_{2d}
ERA	ERS	0.05	50	20	0.800	1.243	0.594	0.146	0.553	0.851	1.387	0.605	0.140	0.520	0.046
ERA	ERS	0.05	50	30	0.841	1.284	0.581	0.151	0.539	0.896	1.433	0.591	0.145	0.505	0.045
ERA	ERS	0.05	60	20	0.929	1.362	0.552	0.162	0.507	0.978	1.500	0.564	0.156	0.476	0.044
ERA	ERS	0.05	60	30	0.982	1.412	0.536	0.169	0.490	1.033	1.554	0.548	0.161	0.460	0.043
ERA	ERS	0.1	50	20	1.307	1.536	0.394	0.295	0.333	1.356	1.683	0.414	0.275	0.300	0.056
ERA	ERS	0.1	50	30	1.412	1.618	0.368	0.314	0.305	1.465	1.770	0.388	0.292	0.275	0.052
ERA	ERS	0.1	60	20	1.467	1.657	0.355	0.325	0.290	1.504	1.789	0.377	0.302	0.263	0.051
ERA	ERS	0.1	60	30	1.590	1.754	0.328	0.348	0.261	1.632	1.889	0.349	0.323	0.236	0.048
ERA	HES	0.05	50	20	3.598	7.554	0.594	0.301	0.554	4.493	10.916	0.646	0.282	0.461	0.149
ERA	HES	0.05	50	30	3.899	8.040	0.581	0.310	0.539	4.827	11.457	0.633	0.291	0.447	0.150
ERA	HES	0.05	60	20	4.553	9.000	0.551	0.335	0.506	5.160	11.827	0.613	0.311	0.420	0.154
ERA	HES	0.05	60	30	4.976	9.648	0.536	0.346	0.490	5.580	12.476	0.598	0.321	0.404	0.154
ERA	HES	0.1	50	20	8.942	14.632	0.394	0.522	0.333	6.785	13.328	0.504	0.474	0.263	0.192
ERA	HES	0.1	50	30	10.233	16.246	0.368	0.549	0.305	7.428	14.150	0.483	0.494	0.242	0.189
ERA	HES	0.1	60	20	10.935	17.085	0.355	0.565	0.290	7.590	14.288	0.474	0.509	0.231	0.191
ERA	HES	0.1	60	30	12.650	19.147	0.328	0.596	0.261	8.368	15.260	0.452	0.532	0.210	0.187
HEA	ERS	0.05	50	20	1.854	3.755	0.702	0.591	0.672	4.103	9.868	0.729	0.527	0.368	0.334
HEA	ERS	0.05	50	30	1.940	3.866	0.695	0.597	0.664	4.411	10.366	0.717	0.534	0.355	0.334
HEA	ERS	0.05	60	20	2.433	4.633	0.664	0.618	0.630	4.898	11.088	0.699	0.544	0.337	0.332
HEA	ERS	0.05	60	30	2.560	4.790	0.655	0.627	0.620	5.305	11.715	0.685	0.552	0.323	0.330
HEA	ERS	0.1	50	20	3.140	5.378	0.589	0.811	0.548	6.310	12.724	0.630	0.707	0.202	0.391
HEA	ERS	0.1	50	30	3.363	5.637	0.576	0.821	0.533	7.104	13.825	0.606	0.718	0.186	0.381
HEA	ERS	0.1	60	20	4.135	6.737	0.546	0.836	0.500	7.416	14.206	0.597	0.723	0.179	0.378
HEA	ERS	0.1	60	30	4.469	7.121	0.530	0.848	0.483	8.441	15.577	0.570	0.736	0.162	0.365
HEA	HES	0.05	50	20	2.724	6.166	0.702	0.584	0.672	6.331	17.011	0.738	0.530	0.396	0.315
HEA	HES	0.05	50	30	2.863	6.385	0.695	0.590	0.664	6.756	17.756	0.728	0.536	0.385	0.316
HEA	HES	0.05	60	20	3.608	7.659	0.664	0.611	0.630	7.351	18.528	0.709	0.549	0.361	0.319
HEA	HES	0.05	60	30	3.818	7.975	0.655	0.619	0.620	7.897	19.437	0.697	0.556	0.348	0.318
HEA	HES	0.1	50	20	5.208	10.094	0.589	0.781	0.548	8.912	20.301	0.654	0.702	0.240	0.379
HEA	HES	0.1	50	30	5.610	10.656	0.576	0.790	0.533	9.735	21.506	0.636	0.710	0.226	0.374
HEA	HES	0.1	60	20	6.863	12.535	0.546	0.806	0.501	10.192	21.985	0.623	0.719	0.212	0.374
HEA	HES	0.1	60	30	7.468	13.348	0.530	0.817	0.483	11.226	23.441	0.604	0.729	0.196	0.368
PCA	ERS	0.05	50	20	0.762	2.337	0.673	0.473	0.640	2.372	12.737	0.682	0.462	0.419	0.230
PCA	ERS	0.05	50	30	0.798	2.407	0.660	0.498	0.626	2.641	13.843	0.669	0.486	0.396	0.239
PCA	ERS	0.05	60	20	0.978	3.094	0.645	0.517	0.610	3.020	15.335	0.658	0.503	0.375	0.248
PCA	ERS	0.05	60	30	1.028	3.195	0.632	0.546	0.595	3.408	16.816	0.643	0.531	0.349	0.257
PCA	ERS	0.1	50	20	1.553	4.613	0.553	0.727	0.508	4.362	19.515	0.574	0.701	0.218	0.311
PCA	ERS	0.1	50	30	1.654	4.825	0.537	0.768	0.490	5.186	22.225	0.555	0.740	0.188	0.320

Table 1. *Cont.*

					Model 1					Model 2					
Arr.	**Ser.**	γ	K	L	μ_1	σ_1	ν_1	ν_1^*	ϑ_1	μ_2	σ_2	ν_2	ν_2^*	ϑ_{2a}	ϑ_{2d}
PCA	ERS	0.1	60	20	2.092	6.237	0.529	0.765	0.481	5.427	23.006	0.555	0.736	0.187	0.321
PCA	ERS	0.1	60	30	2.238	6.543	0.513	0.806	0.464	6.573	26.474	0.536	0.774	0.158	0.328
PCA	HES	0.05	50	20	2.132	5.285	0.673	0.525	0.640	5.142	22.448	0.701	0.496	0.432	0.239
PCA	HES	0.05	50	30	2.290	5.552	0.660	0.549	0.626	5.574	23.685	0.688	0.516	0.413	0.244
PCA	HES	0.05	60	20	2.616	6.316	0.645	0.568	0.610	6.030	24.915	0.678	0.534	0.390	0.256
PCA	HES	0.05	60	30	2.811	6.634	0.632	0.594	0.595	6.589	26.449	0.664	0.556	0.369	0.261
PCA	HES	0.1	50	20	4.248	9.239	0.553	0.762	0.508	7.712	28.419	0.607	0.712	0.254	0.314
PCA	HES	0.1	50	30	4.606	9.777	0.537	0.797	0.490	8.545	30.479	0.590	0.742	0.230	0.319
PCA	HES	0.1	60	20	5.195	11.193	0.529	0.796	0.481	8.870	31.287	0.588	0.742	0.222	0.325
PCA	HES	0.1	60	30	5.625	11.838	0.513	0.830	0.464	9.930	33.825	0.571	0.772	0.199	0.329

Table 2. Inventory subsystem—μ = 1.1, ERS.

						Model 1					Model 2		
Arr.	γ	K	L	$\hat{\mu}_1$	$\hat{\sigma}_1$	Γ_1	ξ_1	κ_1	$\hat{\mu}_2$	$\hat{\sigma}_2$	Γ_2	ξ_2	κ_2
ERA	0.05	50	20	11.692	16.321	47.055	0.735	27.228	12.398	16.642	46.897	0.718	27.861
ERA	0.05	50	30	12.410	16.727	44.536	0.804	24.878	13.155	17.034	44.260	0.789	25.342
ERA	0.05	60	20	15.421	19.918	56.746	0.677	29.537	16.253	20.247	56.578	0.660	30.296
ERA	0.05	60	30	16.403	20.333	54.240	0.735	27.215	17.223	20.632	54.003	0.720	27.791
ERA	0.10	50	20	18.142	17.416	44.610	0.583	17.158	18.849	17.486	44.379	0.566	17.678
ERA	0.10	50	30	19.864	17.752	40.356	0.673	14.855	20.583	17.778	40.015	0.658	15.206
ERA	0.10	60	20	22.995	20.644	54.013	0.515	19.433	23.727	20.715	53.788	0.500	20.003
ERA	0.10	60	30	25.101	20.820	49.822	0.583	17.155	25.839	20.818	49.509	0.569	17.592
HEA	0.05	50	20	22.059	22.423	47.842	0.529	37.781	23.096	21.213	46.314	0.499	40.081
HEA	0.05	50	30	22.786	22.726	46.149	0.563	35.502	23.943	21.331	43.373	0.557	35.894
HEA	0.05	60	20	26.855	26.413	57.532	0.499	40.049	28.058	24.908	55.905	0.462	43.287
HEA	0.05	60	30	27.820	26.714	55.848	0.529	37.782	29.090	24.932	53.056	0.511	39.182
HEA	0.10	50	20	28.909	21.539	46.540	0.376	26.572	28.910	19.603	43.757	0.363	27.541
HEA	0.10	50	30	30.178	21.720	44.134	0.410	24.368	30.311	19.383	39.322	0.431	23.199
HEA	0.10	60	20	34.649	25.150	55.954	0.348	28.752	34.484	22.873	52.992	0.328	30.487
HEA	0.10	60	30	36.260	25.219	53.575	0.376	26.573	36.080	22.419	48.708	0.382	26.208
PCA	0.05	50	20	16.747	17.684	45.521	0.615	32.508	17.322	17.644	45.190	0.603	33.193
PCA	0.05	50	30	18.184	18.117	41.838	0.697	28.701	18.788	18.022	41.335	0.688	29.058
PCA	0.05	60	20	21.269	21.103	55.111	0.553	36.149	21.979	21.000	54.703	0.539	37.141
PCA	0.05	60	30	23.030	21.436	51.562	0.615	32.510	23.778	21.255	50.943	0.604	33.102
PCA	0.10	50	20	23.346	17.188	42.458	0.454	22.055	23.965	16.855	41.808	0.440	22.751
PCA	0.10	50	30	25.980	17.206	36.985	0.541	18.480	26.618	16.736	36.101	0.534	18.762
PCA	0.10	60	20	28.547	20.167	51.733	0.394	25.408	29.262	19.730	50.926	0.379	26.462
PCA	0.10	60	30	31.492	19.905	46.540	0.454	22.055	32.211	19.311	45.446	0.444	22.564

5.1. Impact of Service Rate and Service Time Distribution

In Model 1, customers do not enter the system when inventory level is zero, and when a customer enters the system the inventory level is immediately reduced to reflect the new customer's demand. As a consequence, the service rate and service time distribution do not impact ISS. They do impact ISS in Model 2 because customers who arrive when inventory level is zero wait in line hoping for a replenishment during their wait. However, this impact is minimal because the likelihood of a replenishment during their wait is typically very small.

As it might be expected, the performance measures for CSS improved with increasing service rate in both models. The overall impact of service time variability is to degrade the performance of the system, with HES generating the strongest impact. Specifically, increasing service rate increased the probability of an idle server (ν_1, ν_2) and decreased the number of customers in the system (μ_1, μ_2). Increasing variability of service time distribution, on the other hand, increased the mean (μ_1, μ_2), the standard deviation (σ_1, σ_2), and the coefficient of variation of the number of customers in the system, and increased the percent of time the server is idle with positive inventory (ν_1^*, ν_2^*). The probability of customer loss in Model 1 (ν_1) was unaffected by the service time variability, but ν_2 increased marginally. We also observed that when the service rate is less than the arrival rate, $\nu_1 < \vartheta_1$ and $\nu_2 < \vartheta_2$; and when service rate is greater than the arrival rate, $\nu_1 > \vartheta_1$ and $\nu_2 > \vartheta_2$.

Table 3. Comparing Model 1 and Model 2

Model 1 (μ = 1.1, γ = 0.1, K = 50)												
Arr.	**Ser.**	L	μ_1	σ_1	$\hat{\mu}_1$	$\hat{\sigma}_1$	Γ_1	κ_1	ν_1	ν_1^*	ϑ_1	ϑ_1^*
ERA	ERS	20	1.307	1.536	18.142	17.416	44.610	17.158	0.394	0.295	0.333	0.000
ERA	ERS	30	1.412	1.618	19.864	17.752	40.356	14.855	0.368	0.314	0.305	0.000
ERA	EXS	20	1.844	2.391	18.143	17.415	44.609	17.158	0.394	0.345	0.333	0.000
ERA	EXS	30	2.023	2.566	19.867	17.751	40.353	14.855	0.368	0.366	0.305	0.000
ERA	HES	20	8.942	14.632	18.141	17.415	44.608	17.165	0.394	0.522	0.333	0.000
ERA	HES	30	10.233	16.246	19.865	17.750	40.348	14.861	0.368	0.549	0.305	0.000
HEA	ERS	20	3.140	5.378	28.909	21.539	46.540	26.572	0.589	0.811	0.548	0.000
HEA	ERS	30	3.363	5.637	30.178	21.720	44.134	24.368	0.576	0.821	0.533	0.000
HEA	EXS	20	3.244	5.605	28.909	21.539	46.540	26.570	0.589	0.809	0.548	0.000
HEA	EXS	30	3.476	5.878	30.178	21.720	44.134	24.366	0.576	0.819	0.533	0.000
HEA	HES	20	5.208	10.094	28.909	21.539	46.540	26.573	0.589	0.781	0.548	0.000
HEA	HES	30	5.610	10.656	30.177	21.720	44.134	24.367	0.576	0.790	0.533	0.000
PCA	ERS	20	1.553	4.613	23.346	17.188	42.458	22.055	0.553	0.727	0.508	0.000
PCA	ERS	30	1.654	4.825	25.980	17.206	36.985	18.480	0.537	0.768	0.490	0.000
PCA	EXS	20	1.681	4.771	23.345	17.190	42.460	22.050	0.553	0.732	0.508	0.000
PCA	EXS	30	1.793	4.991	25.982	17.205	36.983	18.479	0.537	0.772	0.490	0.000
PCA	HES	20	4.248	9.239	23.345	17.189	42.458	22.055	0.553	0.762	0.508	0.000
PCA	HES	30	4.606	9.777	25.978	17.207	36.987	18.480	0.537	0.797	0.490	0.000

Table 3. *Cont.*

Model 2 (μ = 1.1, γ = 0.1, K = 50)												
Arr.	**Ser.**	L	μ_2	σ_2	$\hat{\mu}_2$	$\hat{\sigma}_2$	Γ_2	κ_2	ν_2	ν_2^*	ϑ_2	ϑ_2^*
ERA	ERS	20	1.356	1.683	18.849	17.486	44.379	17.678	0.414	0.275	0.356	0.156
ERA	ERS	30	1.465	1.770	20.583	17.778	40.015	15.206	0.388	0.292	0.327	0.160
ERA	EXS	20	1.916	2.678	19.334	17.625	44.312	18.046	0.428	0.314	0.371	0.210
ERA	EXS	30	2.101	2.860	21.112	17.925	39.932	15.509	0.402	0.332	0.342	0.217
ERA	HES	20	6.785	13.328	22.408	18.541	44.228	20.777	0.504	0.474	0.454	0.422
ERA	HES	30	7.428	14.150	24.551	19.020	40.071	18.015	0.483	0.494	0.431	0.439
HEA	ERS	20	6.310	12.724	28.910	19.603	43.757	27.541	0.630	0.707	0.593	0.659
HEA	ERS	30	7.104	13.825	30.311	19.383	39.322	23.199	0.606	0.718	0.567	0.673
HEA	EXS	20	6.480	13.190	29.025	19.635	43.826	27.733	0.632	0.706	0.595	0.654
HEA	EXS	30	7.280	14.306	30.446	19.437	39.457	23.427	0.609	0.717	0.570	0.668
HEA	HES	20	8.912	20.301	30.279	19.907	44.550	30.039	0.654	0.702	0.619	0.612
HEA	HES	30	9.735	21.506	31.912	19.896	40.902	26.200	0.636	0.710	0.600	0.624
PCA	ERS	20	4.362	19.515	23.965	16.855	41.808	22.751	0.574	0.701	0.529	0.587
PCA	ERS	30	5.186	22.225	26.618	16.736	36.101	18.762	0.555	0.740	0.509	0.630
PCA	EXS	20	4.568	20.087	24.067	16.891	41.823	22.856	0.576	0.702	0.533	0.584
PCA	EXS	30	5.392	22.750	26.719	16.774	36.134	18.851	0.557	0.740	0.512	0.625
PCA	HES	20	7.712	28.419	25.607	17.460	42.148	24.912	0.607	0.712	0.568	0.552
PCA	HES	30	8.545	30.479	28.307	17.489	36.858	20.822	0.590	0.742	0.549	0.581

5.2. Effect of Arrival Process

Arrival process impacts both the ISS and the CSS and in general, increasing variability of arrivals had the effect of degrading the performance of the system. It is interesting to note that for systems with both a highly variable arrival process and service time distribution, the combined effect is more attenuated than the individual effects.

Systems with highly variable arrival processes (HEA and PCA relative to ERA) had higher probability of a customer loss (ϑ_1 and ϑ_2), and higher mean (μ_1, μ_2), standard deviation (σ_1, σ_2), and the coefficient of variation of the number of customers in the system. Similarly, HEA and PCA produced higher mean ($\hat{\mu}_1, \hat{\mu}_2$), standard deviation ($\hat{\sigma}_1, \hat{\sigma}_2$) but lower coefficient of variation of the inventory levels. Mean cycle times (ξ_1, ξ_2) were lower but the order quantities (Γ_1, Γ_2) remained relatively unaffected. HEA and PCA increased the probability that the server is idle (ν_1, ν_2) as well as the percent of time the server is idle with inventory of units on hand (ν_1^*, ν_2^*), significantly degrading the system performance. The impact of HEA was more pronounced than PCA.

Figures 4 and 5 summarize the effect of the arrival and service processes on the steady-state marginal distributions of the number of customers in the system. In these figures, X and Y axes are truncated at much smaller values than required by the results in order to highlight the differences in the probability distributions.

Figures 4 and 5 indicate that increasing variability in the system (either for service times or for the arrival process) shifts the probability from states with smaller number in the system to the states with higher number in the system, thus creating a longer tail for the distribution. Model 2, in general, has a slightly longer tail relative to Model 1.

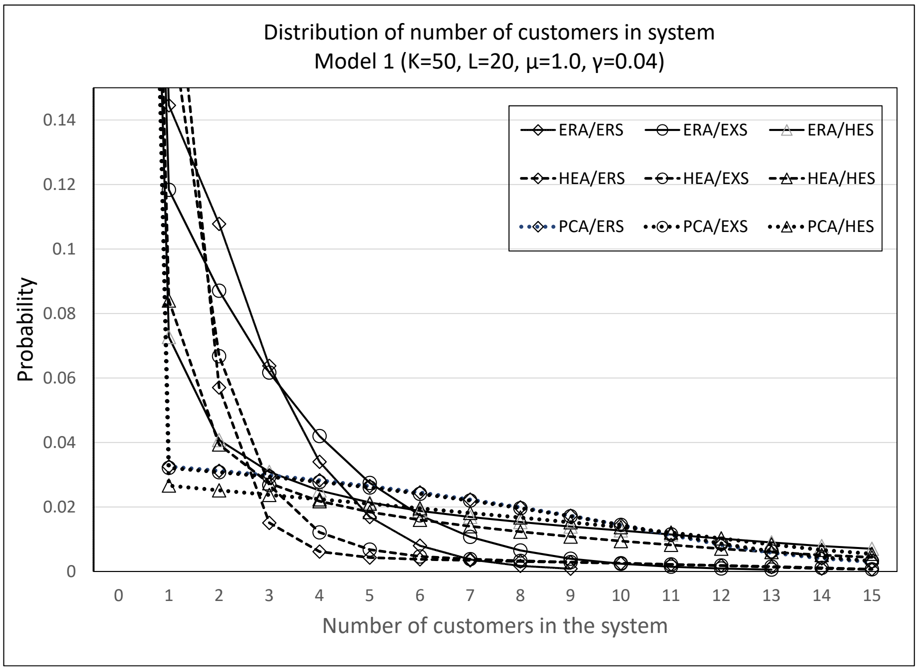

Figure 4. Marginal distributions of number of customers in the system (Model 1).

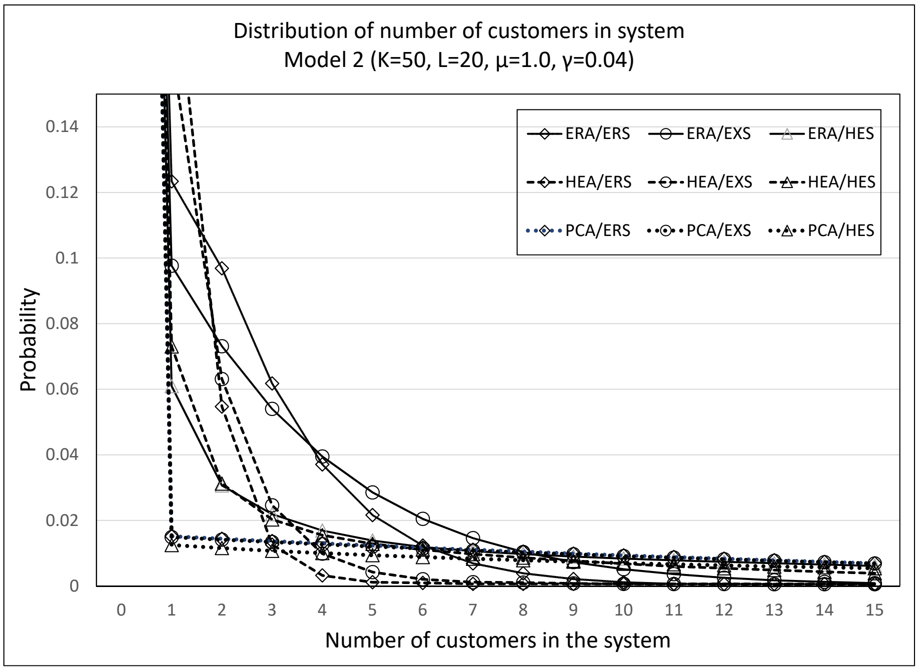

Figure 5. Marginal distributions of number of customers in the system (Model 2).

5.3. Impact of γ

γ is not under the control of the management but understanding its effect on the system is essential to choose K and L efficient operation of the system. We considered values of γ =1/10, 1/15, 1/20, and 1/25 (with $\lambda = 1$) in the numerical analysis.

For a given K and L, larger values of γ resulted in more frequent usage of opportunistic events to procure replenishment or smaller order cycle times, κ_1 and κ_2), and hence smaller size replenishments, Γ_1 and Γ_2, resulting in a net increase in the mean inventory levels $\hat{\mu}_1$ and $\hat{\mu}_2$). This observation should be interpreted with caution because it states the effect of larger γ, keeping K and L at the same level. Larger γ indicates more frequent opportunities for replenishment permitting smaller K and L to achieve the same service level. Larger

γ, in addition to increasing the means, $\hat{\mu}_1$ and $\hat{\mu}_2$, also increased the standard deviations, $\hat{\sigma}_1$ and $\hat{\sigma}_2$, but with a net decrease in the coefficients of variation of the inventory levels. Larger γ also had a similar effect on the numbers of customers in the system by increasing the means, μ_1 and μ_2, and the standard deviations, σ_1 and σ_2, but resulted in a net decrease in the coefficients of variation. Further discussion on the combined effect of changes in γ with other parameters such as K and L is presented in the following.

Inventory system with larger values of γ, indicating frequent replenishment opportunities, will permit maintaining smaller inventories to achieve the same level of service.

5.4. Impact of K and L

In the inventory subsystem, as K increases, with all other parameters remaining the same, more items are procured when opportunistic events occur leading to larger inventories on average. Larger inventories lead to larger cycle times (κ_1 and κ_2) because of an increased likelihood of skipping current opportunities to wait for future ones. The net effect is an increase in the means $\hat{\mu}_1$ and $\hat{\mu}_2$ and also an increase in the standard deviations $\hat{\sigma}_1$ and $\hat{\sigma}_2$, but a reduction in the coefficient of variation of the inventory levels. Changes in L had a slightly different, and smaller, effect on the inventory subsystem than corresponding changes in K. As L increases with all other parameters remaining the same, opportunistic events are availed more often (smaller κ_1 and κ_2) and in smaller quantities. The net result is once again an increase in the means, $\hat{\mu}_1$ and $\hat{\mu}_2$, and an increase in the standard deviations, $\hat{\sigma}_1$ and $\hat{\sigma}_2$), but a reduction in the coefficient of variation, of the numbers in the system.

In the customer subsystem, increasing K or L keeping the rest of the parameters the same, led to an increase in the means, μ_1 and μ_2, and an increase in the standard deviations, σ_1 and σ_2, but a reduction in the coefficient of variation, of the number of customers in the system. Increasing K or L also resulted in a decrease the customer loss probability (ϑ_1, ϑ_2), server idle probability (ν_1, ν_2), and an increase in the percent of server idle time with positive inventory (ν_1^*, ν_2^*). Effect of changes in K are much stronger than corresponding changes in L.

5.5. Cost Analysis

A system that can provide the desired service level at the minimum cost is the goal of any optimization. We consider three costs associated with the system, namely, cost of carrying inventory, cost of placing an order, and cost of lost customers/demand. A fourth item, the cost of keeping customers waiting, is not considered in this study because it was not considered particularly relevant. Let c_1, c_2, and c_3 denote the costs of carrying a unit inventory per unit time, cost of placing an order, and the average cost of a lost customer. Z_1 and Z_2, the total cost per unit time for the two models can be expressed as follows. The system cost per unit time is then given by

$$\begin{aligned} Z_1 &= c_1\,\hat{\mu}_1 + \frac{c_2}{\kappa_1} + c_3\,\vartheta_1 \\ Z_2 &= c_1\,\hat{\mu}_2 + \frac{c_2}{\kappa_2} + c_3\,(\vartheta_{2a} + \vartheta_{2d}) \end{aligned}$$

The first two components in Z_1 and Z_2 represent the cost of the inventory subsystem, and the third item represents the cost of the customer subsystem. The goal of management is to choose the values of K and L that minimize the total cost. Using values of $c_1 = 0.25$, $c_2 = 100$, and $c_3 = 10$, a summary of the cost for a select set of parameter values is presented in Table 4, where the minimum cost values for each combination of K and L are highlighted in bold. In order to find the best combination of values of K and L, a more refined search than in Table 4 needs to be performed.

Table 4. Total cost summary γ = 0.1, μ = 1.1.

Service Dist.		**ERS**			**EXS**			**HES**		
Arr.	***L/K***	**40**	**50**	**60**	**40**	**50**	**60**	**40**	**50**	**60**
					Model 1					
ERA	10	13.085	**12.946**	13.125	13.084	**12.942**	13.123	13.080	**12.941**	13.123
ERA	20	13.982	13.695	13.799	13.981	13.695	13.797	13.979	13.691	13.796
ERA	30	15.304	14.751	14.718	15.303	14.750	14.717	15.301	14.745	14.708
HEA	10	**15.453**	15.999	16.667	**15.453**	16.000	16.669	**15.453**	16.000	16.668
HEA	20	15.925	16.470	17.145	15.926	16.471	17.146	15.926	16.470	17.146
HEA	30	16.444	16.982	17.662	16.445	16.983	17.663	16.444	16.983	17.663
PCA	10	**14.161**	14.445	14.976	**14.164**	14.446	14.981	**14.162**	14.444	14.979
PCA	20	15.355	15.448	15.886	15.356	15.451	15.888	15.355	15.449	15.885
PCA	30	17.125	16.807	17.046	17.127	16.810	17.050	17.125	16.808	17.047
					Model 2					
ERA	10	13.258	**13.184**	13.412	13.363	**13.333**	13.598	**13.913**	14.120	14.596
ERA	20	14.169	13.928	14.077	14.281	14.084	14.260	14.871	14.957	15.364
ERA	30	15.541	14.993	14.979	15.656	15.144	15.161	16.158	16.000	16.289
HEA	10	**15.622**	16.262	17.034	**15.635**	16.283	17.062	**15.778**	16.507	17.358
HEA	20	16.280	16.786	17.465	16.292	16.812	17.502	16.423	17.089	17.901
HEA	30	17.263	17.557	18.106	17.258	17.577	18.143	17.229	17.795	18.550
PCA	10	**14.306**	14.664	15.277	**14.333**	14.703	15.323	**14.618**	15.096	15.803
PCA	20	15.533	15.681	16.174	15.556	15.720	16.228	15.787	16.093	16.710
PCA	30	17.360	17.072	17.347	17.375	17.107	17.401	17.428	17.373	17.817

5.6. Comparing Model 1 and Model 2

Table 3 compares the steady behavior of the system under the two models. For a given set of parameter values, Model 1 performed better than Model 2 in terms of the mean number of customers in the system, mean inventory level, and probability of an idle server. For Model 2, the loss of demand can occur either at arrival, or at service completion epochs. ϑ_2 is always larger than ϑ_1, but the difference is significantly higher for the case of HES. This high loss rate may cause higher idle probability. Higher variability of the arrival process (as in the case of HEA or PSA) appears to mitigate the effect. For Model 2, the proportion of customers lost at departure instants varied from a low of 9.01% to a high of 68.4%. We notice that these percentages are significantly larger for HES, HEA, and PCA.

5.7. Comparing (K, L) System with (s, S) System

In this subsection, we compare the characteristics of the system considered in this paper (the (K, L) system) with the (s, S) system considered in Chakravarthy and Rumyantsev [17]. In the (K, L) system, the replenishment opportunities occur randomly at exponentially distributed intervals with no lead time for delivery. In the (s, S) system in [17], orders are placed as inventory is depleted but the items are received after an exponentially distributed lead time. In other words, the random lead time in the (s, S) system is replaced by the random interval between replenishment opportunities in the $(K - L)$ system. For effective comparison, we use exactly the same parameters for both systems and set $K = S$ and $L = s$, and the mean lead time for delivery in the (s, S) system equal to the mean time between two replenishment opportunities in the (K, L) system. Table 5 presents a summary comparison for γ = 0.1, μ = 1.1, $K = S = 60$, $L = s = 20$ for both Model 1 and Model 2, for all arrival and service distributions considered in Section 3.

The summary indicates that for the combinations of parameters considered, (K, L) system has a smaller server idle probability, smaller customer loss probability but on all other counts (s, S) system had better operational performance for both Model 1 and Model 2. Considering the customer subsystem, (K, L) model has larger mean and standard deviation of the number of customers in the system relative to the (s, S) system. In the inventory subsystem, (K, L) model has a larger cycle time, and a larger mean and standard deviation of the inventory level relative to the (s, S) system. This indicates that orders are placed less often in the (K, L) system, and more units are ordered each time an order is placed, resulting in larger mean inventory level with greater variability. The overall conclusion from these observations is that the additional uncertainty introduced in the system by the random supply process adversely affects the system performance.

Cost computations are not performed because the results depend significantly on the relative values of c_1, c_2, and c_3. Furthermore, direct comparison of total costs could be misleading because one has to optimize the two systems separately and then compare the two optimal solutions to see which performs better. Even that is not very meaningful, because one does not choose between a (K, L) system or an (s, S) system.

Table 5. Comparing the (K, L) system with the (s, S) system.

		(K, L) System (γ = 0.1, μ = 1.1, K = 60, L = 20)							**(s, S) System (γ = 0.1, μ = 1.1, S = 60, s = 20)**						
								Model 1							
Arr.	**Ser.**	ϑ_1	ν_1	μ_1	σ_1	$\hat{\mu}_1$	$\hat{\sigma}_1$	κ_1	ϑ_1	ν_1	μ_2	σ_2	μ_2	$\hat{\sigma}_2$	κ_1
ERA	ERS	0.355	0.290	1.467	1.657	22.995	20.644	19.433	0.423	0.365	1.209	1.475	15.660	16.179	16.221
ERA	EXS	0.355	0.290	2.116	2.644	22.999	20.642	19.439	0.423	0.365	1.686	2.267	15.663	16.178	16.222
ERA	HES	0.355	0.290	10.935	17.085	22.986	20.646	19.437	0.423	0.365	7.784	13.214	15.657	16.179	16.223
HEA	ERS	0.546	0.500	4.135	6.737	34.649	25.150	28.752	0.629	0.592	2.438	4.414	24.481	19.098	25.437
HEA	EXS	0.546	0.501	4.272	7.014	34.650	25.149	28.749	0.629	0.592	2.520	4.601	24.481	19.098	25.436
HEA	HES	0.546	0.501	6.863	12.535	34.649	25.150	28.751	0.629	0.592	4.018	8.259	24.481	19.098	25.437
PCA	ERS	0.529	0.481	2.092	6.237	28.547	20.167	25.408	0.568	0.525	1.220	3.468	21.748	16.972	21.571
PCA	EXS	0.529	0.482	2.239	6.409	28.550	20.163	25.411	0.568	0.525	1.338	3.620	21.749	16.970	21.571
PCA	HES	0.529	0.481	5.195	11.193	28.547	20.163	25.415	0.568	0.525	3.677	7.968	21.747	16.972	21.572
								Model 2							
Arr.	**Ser.**	ϑ_2	ν_2	μ_2	σ_2	$\hat{\mu}_2$	$\hat{\sigma}_2$	κ_2	ϑ_2	ν_2	μ_2	σ_2	μ_2	$\hat{\sigma}_2$	κ_2
ERA	ERS	0.377	0.315	1.504	1.789	23.727	20.715	20.003	0.441	0.385	1.271	1.637	16.357	16.730	16.748
ERA	EXS	0.391	0.330	2.151	2.883	24.273	20.838	20.445	0.453	0.398	1.796	2.600	16.843	16.946	17.113
ERA	HES	0.474	0.421	7.590	14.288	27.673	21.708	23.619	0.524	0.476	6.418	12.959	19.846	18.270	19.642
HEA	ERS	0.597	0.556	7.416	14.206	34.484	22.873	30.487	0.651	0.616	5.766	12.099	25.218	18.109	26.769
HEA	EXS	0.599	0.559	7.594	14.674	34.623	22.909	30.690	0.653	0.618	5.931	12.567	25.307	18.153	26.928
HEA	HES	0.623	0.586	10.192	21.985	36.118	23.210	33.174	0.675	0.643	8.199	19.488	26.326	18.629	28.909
PCA	ERS	0.555	0.508	5.427	23.006	29.262	19.730	26.462	0.583	0.540	4.024	18.494	22.363	17.226	22.283
PCA	EXS	0.557	0.512	5.641	23.527	29.380	19.768	26.565	0.585	0.543	4.220	19.058	22.436	17.207	22.384
PCA	HES	0.588	0.547	8.870	31.287	31.069	20.387	28.799	0.617	0.578	7.246	27.332	23.755	17.800	24.279

6. Concluding Remarks

In this paper, we considered a queuing-inventory system where the replenishment opportunities occur at random intervals at which time the decisions are made with regard to the procurement and the quantities. We proposed an opportunistic system involving two parameters, L and K, similar to the well-known (s, S) inventory management system.

The cut-off point, L, between 0 and K, largely determines whether to avail the opportunistic replenishment with certainty or with a probabilistic rule. The two models considered differ in the way the customers enter the system or are getting lost. In Model 1, the customers are lost when the inventory level is zero (irrespective of whether the server is idle or busy). In Model 2, the customers are allowed to enter even when the inventory level is zero but as long as the server is busy serving. Further, the customers may be lost at a service completion due to lack of inventory to start a new service.

The analysis of the results from an extensive set of numerical experimentation under various scenarios of the parameters indicates that with all other parameters remaining the same, Model 1 appears to have better performance characteristics in terms of the number of customers in the system, inventory levels in the system, and the likelihood of losing a customer.

Variability in the overall system behavior can be due to the variability in the arrival process or the service time distribution. In general, increasing the variability of either the input process or the service time distribution, led to an increase in overall variability in the system behavior. When both arrival process and service time distribution are highly variable, the combined effect is more attenuated than the individual effects.

With the tools developed in this paper and the understanding of the system behavior presented in this paper, it is possible to conduct an efficient search for the value of the key parameters K, and L that minimizes the total system cost. While γ, the rate at which replenishment opportunities arise, is not under the control of the management, one might surmise that an optimal inventory system with larger values of γ will have smaller cost because of the availability of frequent replenishment opportunities making it possible to achieve the same level of service with smaller inventories.

A comparison with the (s, S) system studied by Chakravarthy and Rumyantsev [17] revealed that, the (s, S) system has better operating characteristics in terms of number in the system and inventory levels, apparently due to the random supply process. This could be mitigated by the possibly having a lower purchase cost per unit. This aspect is not studied in this paper. In addition, it would be interesting to compare the system studied here with the (s, S) system in which the inventory level is always brought to level S, irrespective of when the replenishment occurs.

A possible extension to this study could be to investigate systems with both regular channels of purchase (as in the (s, S) system) and the opportunistic replenishment opportunities (as in the (K, L) system) with possibly lower per unit cost of purchase in the context of MAP arrivals of demands and phase-type services.

In this paper, we explored two queuing-inventory models where the replenishment opportunities occur at random. Such models are appropriate for non-critical items (e.g., canned goods, generic medication) where stock outs for short periods of time are not serious. Special replenishment opportunities typically offer lower unit cost with little or no ordering cost, and units are usually available without delay. The results of this paper and the detailed discussion of the system behavior, offer a useful tool to compare the traditional inventory systems (e.g., (s, S) system) with the opportunistic models to determine the most effective method to manage inventories under a more general scenario of using MAP for the demand process and phase-type distributions for the processing of the demands.

Author Contributions: Conceptualization: S.R.C. and B.M.R.; methodology: S.R.C.; software: S.R.C.; validation: S.R.C. and B.M.R.; Formal analysis: S.R.C. and B.M.R.; investigation: S.R.C. and B.M.R.; writing—original draft preparation, S.R.C. and B.M.R.; writing—review and editing, S.R.C. and B.M.R.; visualization, S.R.C. and B.M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: The authors gratefully acknowledge the many suggestions by the anonymous reviewers which significantly improved the presentation of the paper.

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

References

1. Friend, J. K. Stock Control with Random Opportunities for Replenishment. *J. Oper. Res. Soc.* **1960**, *11*, 130–136. [CrossRef]
2. Hurter, A.P.; Kaminsky, F.C. Inventory Control with a Randomly Available Discount Purchase Price. *Oper. Res. Q.* **1968**, *19*, 433–444. [CrossRef]
3. Silver, E.A.; Robb, D.J.; Rahnama, M.R. Random Opportunities for Reduced Cost Replenishments. *IIE Trans.* **1993**, *25*, 111–120. [CrossRef]
4. Gurnani, H. Optimal ordering policies in inventory systems with random demand and random deal offerings. *Eur. J. Oper. Res.* **1996**, *95*, 299–312. [CrossRef]
5. Moinzadeh, K. Replenishment and Stocking Policies for Inventory Systems with Random Deal Offerings. *Manag. Sci.* **1997**, *43*, 334–342. [CrossRef]
6. Feng, Y.; Sun, J. Computing the optimal replenishment policy for inventory systems with random discount opportunities. *Oper. Res.* **2001**, *49*, 790–795. [CrossRef]
7. Goh, M.; Sharafali, M. Price-dependent inventory models with discount offers at random times. *Prod. Oper. Manag.* **2002**, *11*, 139–156. [CrossRef]
8. Chaouch, B.A. Inventory control and periodic price discounting campaigns. *Nav. Res. Logist.* **2007**, *54*, 94–108. [CrossRef]
9. Tajbakhsh, M.M.; Lee, C.-G.; Zolfaghari, S. An inventory model with random discount offerings. *Omega* **2011**, *39*, 710–718. [CrossRef]
10. Karimi-Nasab, M.; Konstantaras, I. An inventory control model with stochastic review interval and special sale offer. *Eur. J. Oper. Res.* **2013**, *227*, 81–87. [CrossRef]
11. Den Boer, A.; Perry, O.; Zwart, B. Dynamic pricing policies for an inventory model with random windows of opportunities. *Nav. Res. Logist.* **2018**, *65*, 660–675.
12. Krishnamoorthy, A.; Shajin, D.; Narayanan, V. Inventory with positive service time: A survey. In *Advanced Trends in Queueing Theory*; Mathematics and Statistics Series; ISTE & Wiley: London, UK, 2019. [CrossRef]
13. Choi, K.H.; Yoon, B.K.; Moon, S.A. Queueing Inventory Systems with Phase-type Service Distributions: A Literature Review. *Ind. Eng. Manag. Syst.* **2019**, *18*, 330–339. [CrossRef]
14. Melikova, A. Z.; Shahmaliyeva, M. O. Markov Models of Inventory Management Systems with a Positive Service Time. *J. Comput. Syst. Sci. Int.* **2018**, *57*, 766–783. [CrossRef]
15. Chakravarthy, S.R.; Maity, A.; Gupta, U.C. Modeling and Analysis of Bulk Service Queues with an Inventory under (s, S) Policy. *Ann. Operat. Res.* **2017**, *258*, 263–283.
16. Chakravarthy, S.R. Queueing-inventory models with batch demands and positive service times. *Math. Game Theory Appl.* **2019**, *11*, 95–120. [CrossRef]
17. Chakravarthy, S.R.; Rumyantsev, A. Analytical and simulation studies of queueing-inventory models with MAP demands in batches and positive phase type services. *Simul. Model. Pract. Theory* **2020**, *103*, 102092.
18. Alfa, A.S. *Queueing Theory for Telecommunications*; Springer Science+Business Media, LLC: Berlin, Germany, 2010.
19. Artalejo, J.R.; Gomez-Correl, A.; He, Q.M. Markovian arrivals in stochastic modelling: A survey and some new results. *SORT* **2010**, *34*, 101–144.
20. Chakravarthy, S.R. The batch Markovian arrival process: A review and future work. In *Advances in Probability Theory and Stochastic Processes*; Krishnamoorthy, A., Ed.; Notable Publications Inc.: Trenton, NJ, USA, 2001; pp. 21–39.
21. Chakravarthy, S.R. Markovian Arrival Processes. In *Wiley Encyclopedia of Operations Research and Management Science, John Wiley & Sons, New York, NY, USA*; 2010.
22. Chakravarthy, S.R. Matrix-Analytic Queueing Models. In *An Introduction to Queueing Theory*, 2nd ed.; Narayan Bhat, U., Ed.; Birkhauser; Springer Science + Business Media: New York, NY, USA, 2015; Chapter 8.
23. He, Q.-M. *Fundamentals of Matrix-Analytic Methods*; Springer: New York, NY, USA, 2014. [CrossRef]
24. Lucantoni, D.M.; Meier-Hellstern, K.S.; Neuts, M.F. A single-server queue with server vacations and a class of nonrenewal arrival processes. *Adv. Appl. Probl.* **1990**, *22*, 676–705. [CrossRef]
25. Lucantoni, D.M. New results on the single server queue with a batch Markovian arrival process. *Stoch. Model.* **1991**, *7*, 1–46. [CrossRef]
26. Neuts, M.F. A versatile Markovian point process. *J. Appl. Prob.* **1979**, *16*, 764–779.
27. Neuts, M.F. *Structured Stochastic Matrices of $M/G/1$ Type and Their Applications*; Marcel Dekker, Inc.: New York, NY, USA, 1989.
28. Neuts, M.F. Models based on the Markovian arrival process. *IEICE Trans. Commun.* **1992**, *E75B*, 1255–1265.
29. Neuts, M.F. *Algorithmic Probability: A Collection of Problems*; Chapman and Hall: New York, NY, USA, 1995.
30. Neuts, M.F. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*; The Johns Hopkins University Press: Baltimore, MD, USA, 1981. [1994 version is Dover Edition].
31. Neuts, M.F. Probability distributions of phase type. In *Liber Amicorum Prof. Emeritus H. Florin*; Department of Mathematics, University of Louvain: Ottignies-Louvain-la-Neuve, Belgium, 1975; pp. 173–206.
32. Marcus, M.; Minc, H. *A Survey of Matrix Theory and Matrix Inequalities*; Courier Corporation: Chelmsford, MA, USA, 1992.

33. Steeb, W.-H.; Hardy, Y. *Matrix Calculus and Kronecker Product: A Practical Approach to Linear and Multilinear Algebra*, 2nd ed.; World Scientific Publishing Company: Singapore, 2011.
34. Latouche, G.; Ramaswami, V. *Introduction to Matrix Analytic Methods in Stochastic Modeling*; ASA-SIAM: Philadelphia, PA, USA, 1999.