Queueing Models and
Service Management
Vol. 3, No. 1, page 1-35, 2020

Queueing-Inventory Models for a Two-Vendor System with Positive Service Times

Srinivas R. Chakravarthy[1,*] and Khizer Hayat[2]

[1]Departments of Industrial and Manufacturing Engineering & Mathematics
Kettering University,Flint, MI-48504, USA
[2]Department of Engineering, IFIVEO Inc., San Francisco, CA-94105, USA.

(Received May 2019 ; accepted October 2019)

Abstract: Queueing-inventory systems wherein the demands are processed with random service times have been getting a lot of attention recently. In such systems, each demand needs one or more inventory items and needs positive processing times. The inventory is generally replenished using (s, S)-type policy and the lead times are assumed to be random. In this paper we introduce the concept of multiple vendors who will be responsible for the replenishment of inventory. Under the assumptions of a two-vendor system wherein the demands occurring according to a Markovian arrival process, the service times to be of phase type, and the lead times to be exponentially distributed with parameter depending on the vendor, we analyze the model in steady-state using the well-known matrix-analytic methods. Some interesting numerical examples are presented including one comparing the one and the two-vendor systems.

Keywords: Algorithmic probability, Markovian arrival process, phase type distribution, queueing-inventory systems, two vendors.

1. Introduction and Motivation

It is well-known that service industries use queueing models to figure out the resource allocation as well as to improve their quality of service. Generally speaking, customers requiring services need some form of inventory for servicing them or their products. First known study wherein the inventory is incorporated into queueing systems is due to Bradley [5]. In this context, we also refer to Saffari *et al.* [30], Schwarz *et al.* [31], and Schwarz and Daduna [32]. In classical inventory systems (Nahmias [23]) the customers' demands for inventory are met instantaneously. That is, the demands for the inventory are met with no significant service times. The first paper to introduce a significant service time (also referred to as a positive service time) in processing the inventory is that of Berman *et al.* [3]. Since then the literature on queueing-inventory systems has grown significantly and we refer the reader to the survey papers (see Choi and Yoon [11], Krishnamoorthy *et al.* [17]) as well as to the papers (see Benny *et al.* [2], Chakravarthy *et al.* [10], Krishnamoorthy *et al.* [18], Yue and Qin [39]) and the references therein.

*Corresponding author
Email : schakrav@kettering.edu

While the literature on queueing-inventory systems contains a variety of models dealing with positive service times, one common thing with those papers dealing with (s, S) -type inventory policy is that there is only one source for replenishment. However, in many applications, one deals with multiple vendors.

The motivation for the study of a multiple-vendor problem arose out of a practical need in many areas including production and manufacturing. In industries such as the automotive industry a dual or multi-supplier part-sourcing strategy is used by organizations to deal with supply chain risks. Supply chain disruptions and risks can be caused by a number of sources which may include natural disasters like the Kobe earthquake, SARS, foot and mouth disease, birds flu, terrorist attacks such as $9/11$, industrial or direct action such the fuel price protest in the UK in September 2000 etc. (Cranfield Management School [13]). The primary purpose of this strategy is to make sure that the parts are delivered on time in order to minimize production delays due to shortage of parts and minimize revenue loss. By using this strategy the risk is spread across multiple vendors hence the probability of on-time delivery is significantly increased.

A great example of the effectiveness of having multi-vendors for replenishing inventory is the fire at Toyota's sole supplier of brake-fluid proportioning valves, Aisin, in Reitman [29] that could have caused their production line to shut down for weeks, but their operations kept going. 99% of Toyota's P-valves were made at Aisin with 1% built at Nisshin Kogyo Co. It was estimated that each day Toyota production was halted would lead to a 0.1% decrease in Japan's industrial output. Aisin, along with Toyota, set up a crisis room to deal with the problem of manufacturing new P-valves. Toyota managed to get many of its suppliers to bring in additional engineers, and work overtime shifts, to help build machines to produce P-valves, as well as increase production of the components. While observers initially predicted that Toyota would have to halt production for weeks, the incident ultimately set Toyota's production back only five days.

The Taiwan earthquake of September 1999 is a counter example of when a single vendor (for replenishment) strategy cost electronics firms billions of dollars because their sole suppliers were Taiwanese manufacturers (Sheffi [33]). Another interesting counter example is of the March 2000 fire at the Philips microchip plant in Albuquerque, NM when a lightning bolt struck a power line causing a fire in a production room at Philips. This caused major disruptions for their customers Nokia and Ericsson. Nokia was able to battle this disruption using their multi-tiered supplier strategy to source chips from other suppliers. However, Ericsson could not avoid a production shut down as they were using a single vendor to supply the microchips from the Albuquerque, NM plant causing them to suffer \$400 million in lost sales and it has also withdrawn completely from the mobile phone handset production business (Latour [20]). The case of Ericsson is a more extreme example of the effects of supply chain disruptions but by no means is it exaggerated since these phenomena are experienced by businesses the world over.

The examples above show how a multi-tiered replenishing strategy has benefited organizations and saved them millions to billions of dollars in potential lost revenue due to unexpected supply breakdown. Scholarly articles supporting this point of view include the works of Wong *et al.* [37], who have found that today's supply chains are built to be lean and efficient, but if they are unable to find alternatives quickly for unexpected disruptions, the chains will be susceptible to system shocks and disruptions leading to lost revenue opportunity. Further studies by Christopher and Towill [12] and Tang [36] suggest that as many firms implement various initiatives such as lean, agile, outsourcing, customized, and global networks to gain cost advantage and market share, their supply chains become more vulnerable at the same time, because there tends to be very little inventory in the system to "buffer" any interruptions in supply. As a result, any disruption can have a dramatic impact on the entire chain (Yu *et al.* [38]).

There has been scholarly work advocating a dual-sourcing strategy. Prior work in this area includes a 2008 publication by Yu et al. where they evaluate the impact of supply disruptions risks on the choice between single and dual sourcing methods in a two-stage supply chain with a non-stationary and price-sensitive demand. A critical factor included in their research was two suppliers located in different countries where supplier 1 is located outside the manufacturer's country, offers competitive price but is more prone to breakdowns whereas supplier 2 is local, stable but more expensive (Yu *et al.* [38]).

This phenomenon is what the modern supply chains closely resemble as a lot of Fortune 500 manufacturers are beginning to source parts from countries with cheaper labor costs such as China and Mexico where the probability of delayed deliveries is higher. In order to battle that issue, they keep a secondary supplier close to their manufacturing facilities either within the same state or a nearby state to reduce the risk of delayed deliveries by splitting the sourcing of the parts.

Their paper concluded that there are two critical values of the disruption probabilities which provide a guideline for choosing the most profitable sourcing method for the buying firms. They found that either single or dual sourcing can be effective depending on the magnitude of the disruption probability. A disruption probability of less than 0.08 would lead to a better single sourcing strategy with just the main international supplier. However, a disruption probability between 0.08 and 0.25 would lead to dual sourcing as the best strategy, and a disruption probability of greater than 0.25 would lead to shifting all parts to a local supplier. This model does have certain limitations such as considering a fairly simple demand model without time-related factors as well as considering supplier capacity to be infinite when in reality the capacity is always finite (Yu *et al.* [38]). However, a disruption probability between 0.08 and 0.25 is not too far from what is observed in the real-world as seen from the examples mentioned above. Adding finite capacity would increase the disruption probabilities of the suppliers as well hence the range of 0.08 and 0.25 is a good reference point for practical use.

Therefore, in modern industry a dual or multi-sourcing strategy is widely adopted as the advantages outweigh the disadvantages. This is primarily because new technologies and products are much more complex than in the past and the cost of switching suppliers is relatively low compared to the benefits of reduced probability of supply chain disruptions. Furthermore, apart from the fact that a multi-sourcing strategy lowers the risk associated with supply chain disruption, it also increases competition amongst suppliers resulting in better quality of products with higher reliability.

In order to set up the needed matrices to form the generator of the Markov process describing the model under study, we require the following notation.

- $\boldsymbol{e}$ is a column vector (of appropriate dimension) of 1's.
- $\boldsymbol{e}_i$ is a unit column vector (of appropriate dimension) with 1 in the i^{th} position and 0 elsewhere.
- I is an identity matrix (of appropriate dimension).
- Generally the dimension of the vectors and matrices should be clear in the context of usage. However, when more clarity is needed we will include the dimension. For example, we will use $e(mn)$ to show that the column vector is of dimension mn.
- The notation $'$ stands for the transpose of a matrix or a vector.
- The symbols, $\otimes$ and $\oplus$, respectively, will stand for the Kronecker product and Kronecker sum of matrices. We refer the reader to Graham [14], Marcus and Minc [22], and Steeb [34] for details and properties on Kronecker products.

The paper is organized as follows. In Section 2, we describe the queueing-inventory system under study in this paper and the steady-state analysis of the model is presented in Section 3. Illustrative numerical examples are presented in Section 4 and some concluding remarks including future research work are presented in Section 5.

2. Model Assumptions

In this section, we will describe the model under study. The demands arrive according to a point process. Normally the demands arrive from different sources and hence it is possible for two successive inter-demand times to be dependent. That is, there is a possibility of correlation (positive or negative) present in the inter-demand times. This is very common, especially, when the demands occur from different sources and the sources are not modeled using Poisson processes. A most widely used and popular point process that will suit modeling such inter-demand times is the Markovian arrival process (MAP), a special case of batch Markovian arrival process $(BMAP)$, introduced first by Neuts [25] as a Versatile Markovian Point Process $(VMPP)$. The MAP is represented by a pair of square matrices, say, D_0 and D_1, of order m such that $D = D_0 + D_1$ is an irreducible generator. The entries of D_0 govern transitions corresponding to no arrivals of demands and the entries of D_1 govern transitions governing the arrivals of demands to the system. Suppose that η is the steady-state probability vector of the generator D. That is, $\boldsymbol{\eta}$ satisfies

$$\boldsymbol{\eta} D = 0, \quad \boldsymbol{\eta} e = 1. \tag{1}$$

The constant $\lambda = \boldsymbol{\eta} D_1 \boldsymbol{e}$, known as the fundamental rate, gives the expected number of arrivals per unit of time in the stationary version of MAP.

An arriving demand finding the inventory level to be zero will be considered lost; however, if an arriving demand finds the inventory level to be positive will enter into the system. At that time the inventory level will be reduced by one and the admitted demand will either enter into service immediately provided the server is idle or wait in the queue of infinite size and be served on a first-come-first-served $(FCFS)$ basis.

We assume that the service times are of phase type with representation $(\boldsymbol{\beta}, T)$ of order n. Note that phase type distributions ($PH-$distributions), introduced by Neuts [24] and studied extensively by Neuts and his colleagues (see e.g. Bladt and Nielsen [4], Buchholz *et al.* [6], Chakravarthy [7, 8, 9], He [15], Lucantoni et al. [21], and Neuts [25, 26, 27, 28]) generalize some well-known distributions like exponential, Erlang and hyperexponential among many others.

Suppose that μ is the service rate. It can be seen that $\mu = [\boldsymbol{\beta}(-T)^{-1}\boldsymbol{e}]^{-1}$. Let $\boldsymbol{\zeta}$ denote the stationary probability vector of $T + \boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}$. That is, $\boldsymbol{\zeta}$ satisfies

$$\boldsymbol{\zeta}(T + \boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}) = \mathbf{0}, \quad \boldsymbol{\zeta} \boldsymbol{e} = 1. \tag{2}$$

It can readily be seen (see e.g., [Neuts [26]) that $\boldsymbol{\zeta}$ is given explicitly as

$$\boldsymbol{\zeta} = \mu \boldsymbol{\beta}(-T)^{-1}. \tag{3}$$

We adapt the classical (s, S)-policy for replenishment. That is, in the classical setting, whenever the inventory level drops at or below s, an order for $S - s$ items is placed to replenish the inventory so that the at the time of the arrival of the order the inventory level can be brought to a level in the interval $[S - s, S]$. Also, in the classical setting, it is assumed that $S - s > s$ to avoid placing an order for replenishment at the time of the delivery of a replenishment.

In this paper, we assume that there are two vendors who will be responsible for replenishing the inventory. Vendor 1 is responsible to replenish S_1 items when an order is placed with them, and Vendor 2 is responsible for S_2 items when an order is placed with them. Note that $S_1 + S_2 = S - s$. Without loss of generality we assume that $S_1 \geq S_2$ and further we will assume that $S-s>s$. When the inventory level drops to s or below, one of the following events will occur: (a) if no replenishment is pending, orders are placed with both the vendors; (b) if replenishment is pending from both the vendors, no further order is placed; (c) if only one vendor's order is pending, order is placed only with the other vendor. That is, if Vendor 1 (or Vendor 2) is pending, then an order for S_2 (or S_1) items is placed with Vendor 2 (or Vendor 1). Thus, at any given time there can at most be one replenishment pending with any vendor. Note that it is possible for both vendors to have exactly one pending order. This is similar to the classical inventory system with one vendor doing the replenishment and at most one order can be pending at any given time.

Finally, we assume that the lead times for replenishment for Vendors 1 and 2 are exponential with parameter, respectively, with parameters θ_1 and θ_2.

3. Steady-State Analysis

In this section we will discuss the steady-state analysis of the model described in 2. Towards this end, we split this section into several subsections to lay the ground work.

3.1. The QBD process

The model described in Section 2 can be studied as a $QBD-$process. To describe the state space of the process, we first define $N_1(t)$ to be the number of customers in the system; $N_2(t)$ to be the level of the inventory; $J_1(t)$ to be the status of the replenishment; $J_2(t)$ to be the phase of the service time; and $J_3(t)$ to be the phase of the arrival process at time t. Note that the when the server is idle, $J_2(t)$ will be undefined. $J_1(t)$ is defined as follows.

$$J_1(t)=\begin{cases}0, & \textit{when no replenishment is pending},\\ 1, & \textit{when replenishment is pending only from Vendor } 1,\\ 2, & \textit{when replenishment is pending only from Vendor } 2,\\ 3, & \textit{when replenishment is pending from both Vendors}.\end{cases} \tag{4}$$

First observe that whenever the inventory level is in the set $\{0,1,\cdots,s\}$, orders for replenishment will be pending from both vendors. This could happen in one of three ways. As soon as the inventory level hits s (i) orders are placed with both the vendors since no order is pending from either of the vendors; (ii) an order for S_1 items is placed with Vendor 1 since Vendor 2 order is still pending; and (iii) an order for S_2 items is placed with Vendor 2 since Vendor 1 order is still pending. Secondly, at the time when the inventory is replenished from Vendor 1 while Vendor 2 order is still pending, the inventory level will belong to the set $\{S_1,\cdots,s+S_1\}$. However, during the time when an order is pending from Vendor 2, the inventory level go down from $s+S_1$ to $s+1$. In a similar manner, the possible inventory levels when an order is pending only from Vendor 1 are $\{s+1,\cdots,s+S_2\}$. These observations will help us to define the state space of the system.

It is easy to verify that $\{(N_1(t), N_2(t), J_1(t), J_2(t), J_3(t)) : t \geq 0\}$ is a quasi-birth-and-death process (QBD) with state space given by

$$\begin{aligned}\Omega = &\{(0, j, 3, k_3) : 0 \leq j \leq s, 1 \leq k_3 \leq m\}\\ &\bigcup\{(0, j, 0, k_3) : s+1 \leq j \leq S, 1 \leq k_3 \leq m\}\\ &\bigcup\{(0, j, 1, k_3) : s+1 \leq j \leq s+S_2, 1 \leq k_3 \leq m\}\\ &\bigcup\{(0, j, 2, k_3) : s+1 \leq j \leq s+S_1, 1 \leq k_3 \leq m\}\\ &\bigcup\{(i, j, 3, k_2, k_3) : i \geq 1, 0 \leq j \leq s, 1 \leq k_2 \leq n, 1 \leq k_3 \leq m\}\\ &\bigcup\{(i, j, 0, k_2, k_3) : i \geq 1, s+1 \leq j \leq S, 1 \leq k_2 \leq n, 1 \leq k_3 \leq m\}\\ &\bigcup\{(i, j, 1, k_2, k_3) : i \geq 1, s+1 \leq j \leq s+S_2, 1 \leq k_2 \leq n, 1 \leq k_3 \leq m\}\\ &\bigcup\{(i, j, 2, k_2, k_3) : i \geq 1, s+1 \leq j \leq s+S_1, 1 \leq k_2 \leq n, 1 \leq k_3 \leq m\}.\end{aligned}$$

Let $\mathbf{0} = \{(0, j, 3, k_3), 0 \leq j \leq s, 1 \leq k_3 \leq m\} \bigcup \{(0, j, 0, k_3), s+1 \leq j \leq S, 1 \leq k_3 \leq m\} \bigcup \{(0, j, 1, k_3) : s+1 \leq j \leq s+S_2, 1 \leq k_3 \leq m\} \bigcup \{(0, j, 2, k_3) : s+1 \leq j \leq s+S_1, 1 \leq k_3 \leq m\}$, $1 \leq k_3 \leq m$ denote the set of states corresponding to the system wherein the server is idle; let $\boldsymbol{i} = \{(i, j, 3, k_2, k_3), 0 \leq j \leq s, 1 \leq k_2 \leq n, 1 \leq k_3 \leq m\} \bigcup \{(i, j, 0, k_2, k_3), s+1 \leq j \leq S, k_3 \leq 1 \leq k_2 \leq n, 1 \leq m\} \bigcup \{(i, j, 1, k_2, k_3) : s+1 \leq j \leq s+S_2, 1 \leq k_2 \leq n, 1 \leq k_3 \leq m\} \bigcup \{(0, j, 2, k_2, k_3) : s+1 \leq j \leq s+S_1, 1 \leq k_2 \leq n, 1 \leq k_3 \leq m\}$, denote the set of states wherein exactly i customers are in the system with the inventory level (not including the one the customers already possess and waiting for service), the phases of the service and arrivals in their respective phases.

The generator, Q, of the QBD process under consideration is of the form

$$Q = \begin{bmatrix} B_1 & B_0 & & & \\ B_2 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & A_2 & A_1 & A_0 \\ & & \ddots & \ddots & \ddots \\ & & & & \end{bmatrix}, \tag{5}$$

where the (block) matrices appearing in Q are as follows. In the following we denote $\theta = \theta_1 + \theta_2$ and a diagonal (block) matrix with elements $C_1, \cdots, C_r$ by $\Delta(C_1, \cdots, C_r)$.

$$B_1 = \begin{bmatrix} B_{11}^{(1)} & B_{12}^{(1)} & B_{13}^{(1)} \\ B_{21}^{(1)} & B_{22}^{(1)} & 0 \\ B_{31}^{(1)} & 0 & B_{33}^{(1)} \end{bmatrix}, \quad B_{12}^{(1)} = \begin{bmatrix} 0 & \theta_2 I_{s+1} \\ 0 & 0 \end{bmatrix}, \quad B_{13}^{(1)} = \begin{bmatrix} 0 & \theta_1 I_{s+1} \\ 0 & 0 \end{bmatrix},$$

$$B_{11}^{(1)} = \Delta(D - \theta I, D_0 - \theta I, \cdots, D_0 - \theta I, D_0, \cdots, D_0), \; B_{21}^{(1)} = \begin{bmatrix} 0 & \theta_1 I_{S_2} \end{bmatrix}, \tag{6}$$

$$B_{22}^{(1)} = I_{S_2} \otimes (D_0 - \theta_1 I), \; B_{31}^{(1)} = \begin{bmatrix} 0 & \theta_2 I_{S_1} \end{bmatrix}, \; B_{33}^{(1)} = I_{S_1} \otimes (D_0 - \theta_2 I),$$

$$B_2 = (I \otimes \boldsymbol{T}^{\mathbf{0}} \otimes I),\ B_0 = \begin{bmatrix} B_{11}^{(0)} & 0 & 0 \\ B_{21}^{(0)} & B_{22}^{(0)} & 0 \\ B_{31}^{(0)} & 0 & B_{33}^{(0)} \end{bmatrix},\ B_{11}^{(0)} = \begin{bmatrix} 0 & 0 \\ I_S \otimes \boldsymbol{\beta} \otimes D_1 & 0 \end{bmatrix},$$

$$B_{21}^{(0)} = \boldsymbol{e}'_{s+1}(S+1) \otimes \boldsymbol{e}_1(S_2) \otimes \boldsymbol{\beta} \otimes D_1,\ B_{22}^{(0)} = \begin{bmatrix} 0 & 0 \\ I_{S_2-1} \otimes \boldsymbol{\beta} \otimes D_1 & 0 \end{bmatrix}, \tag{7}$$

$$B_{31}^{(0)} = \boldsymbol{e}'_{s+1}(S+1) \otimes \boldsymbol{e}_1(S_1) \otimes \boldsymbol{\beta} \otimes D_1,\ B_{33}^{(0)} = \begin{bmatrix} 0 & 0 \\ I_{S_1-1} \otimes \boldsymbol{\beta} \otimes D_1 & 0 \end{bmatrix},$$

$$A_0 = \begin{bmatrix} A_{11}^{(0)} & 0 & 0 \\ A_{21}^{(0)} & A_{22}^{(0)} & 0 \\ A_{31}^{(0)} & 0 & A_{33}^{(0)} \end{bmatrix},\ A_{11}^{(0)} = \begin{bmatrix} 0 & 0 \\ I_{Sn} \otimes D_1 & 0 \end{bmatrix},$$

$$A_{21}^{(0)} = \boldsymbol{e}'_{s+1}(S+1) \otimes \boldsymbol{e}_1(S_2) \otimes I \otimes D_1,\ A_{22}^{(0)} = \begin{bmatrix} 0 & 0 \\ I_{(S_2-1)n} \otimes D_1 & 0 \end{bmatrix}, \tag{8}$$

$$A_{31}^{(0)} = \boldsymbol{e}'_{s+1}(S+1) \otimes \boldsymbol{e}_1(S_1) \otimes I \otimes D_1,\ A_{33}^{(0)} = \begin{bmatrix} 0 & 0 \\ I_{(S_1-1)n} \otimes D_1 & 0 \end{bmatrix},$$

$$A_1 = \begin{bmatrix} A_{11}^{(1)} & A_{12}^{(1)} & A_{13}^{(1)} \\ A_{21}^{(1)} & A_{22}^{(1)} & 0 \\ A_{31}^{(1)} & 0 & A_{33}^{(1)} \end{bmatrix},\ A_{12}^{(1)} = \begin{bmatrix} 0 & \theta_2 I_{s+1} \\ 0 & 0 \end{bmatrix},\ A_{13}^{(1)} = \begin{bmatrix} 0 & \theta_1 I_{s+1} \\ 0 & 0 \end{bmatrix},$$

$$A_{11}^{(1)} = \Delta(T \oplus D - \theta I, T \oplus D_0 - \theta I, \cdots, T \oplus D_0 - \theta I, T \oplus D_0, \cdots, T \oplus D_0),$$

$$A_{21}^{(1)} = \begin{bmatrix} 0 & \theta_1 I_{S_2} \end{bmatrix},\ A_{22}^{(1)} = I_{S_2} \otimes (T \oplus D_0 - \theta_1 I),\ A_{31}^{(1)} = \begin{bmatrix} 0 & \theta_2 I_{S_1} \end{bmatrix}, \tag{9}$$

$$A_{33}^{(1)} = I_{S_1} \otimes (T \oplus D_0 - \theta_2 I),\ A_2 = I_{2S-s+1} \otimes \boldsymbol{T}^{\mathbf{0}} \boldsymbol{\beta} \otimes I.$$

The transition diagram at a somewhat macro level is displayed in Figure 1 below. Note that in the diagram the set of states, the following notations are used.

- $\mathbf{(0,3)} = \{(i, j, r, k_3), 0 \le j \le s, 1 \le k_3 \le m\}$
- $\mathbf{(0,0)} = \{(0, j, 0, k_3), s+1 \le j \le S, 1 \le k_3 \le m\}$
- $\mathbf{(0,1)} = \bigcup \{(0, j, 1, k_3) : s+1 \le j \le s+S_2, 1 \le k_3 \le m\}$
- $\mathbf{(0,2)} = \bigcup \{(0, j, 2, k_3) : s+1 \le j \le s+S_1, 1 \le k_3 \le m\}$
- $(\boldsymbol{i},\mathbf{3}) = \{(i, j, 3, k_2, k_3), 0 \le j \le s, 1 \le k_2 \le n, 1 \le k_3 \le m\}, i \ge 1$
- $(\boldsymbol{i},\mathbf{0}) = \{(i, j, 0, k_2, k_3), s+1 \le j \le S, 1 \le k_2 \le n, 1 \le k_3 \le m\}, i \ge 1$
- $(\boldsymbol{i},\mathbf{1}) = \{(i, j, 1, k_2, k_3) : s+1 \le j \le s+S_2, 1 \le k_2 \le n, 1 \le k_3 \le m\}, i \ge 1$

- $(\boldsymbol{i},\mathbf{2}) = \{(0, j, 2, k_2, k_3) : s+1 \le j \le s+S_1, 1 \le k_2 \le n, 1 \le k_3 \le m\}, i \ge 1.$

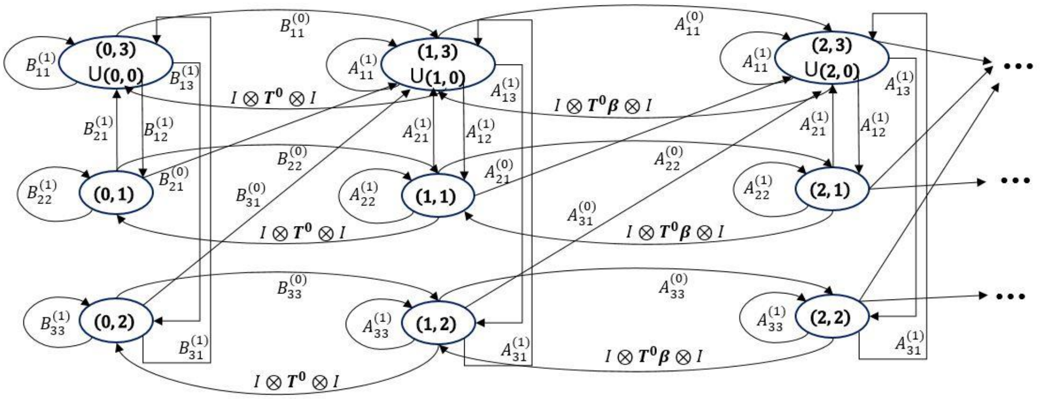

Figure 1. Transition diagram at semi-macro level.

Suppose that $\boldsymbol{\pi} = (\pi_0, \cdots, \pi_{2S-s})$ denotes the steady-state probability vector of $A = A_0 + A_1 + A_2$. That is,

$$\boldsymbol{\pi} A = \mathbf{0}, \quad \boldsymbol{\pi e} = 1. \tag{10}$$

The following theorem gives the stability condition of the queueing-inventory model under study.

Theorem 1. *The queueing-inventory model under study with the generator given in (5) is stable if and only if*

$$[\lambda - \pi_0(\boldsymbol{e} \otimes D_1 \boldsymbol{e})] < \mu. \tag{11}$$

Proof. The proof follows immediately on applying the stability condition for the $QBD-$ process, which is $\boldsymbol{\pi} A_0 \boldsymbol{e} < \boldsymbol{\pi} A_2 \boldsymbol{e}$ (see, e.g., [Neuts [26]).

Note: The stability condition will become obvious once we show that $[\lambda - \pi_0(\boldsymbol{e} \otimes D_1 \boldsymbol{e})]$ is the effective arrival rate as the quantity $\pi_0(\boldsymbol{e} \otimes D_1 \boldsymbol{e}) / \lambda$ gives the probability that an arrival will be lost due to zero inventory.

In the sequel, we define the traffic intensity, ρ, of the model under study as

$$\rho = \frac{\lambda - \pi_0(\boldsymbol{e} \otimes D_1 \boldsymbol{e})}{\mu}. \tag{12}$$

The following lemma is intuitively obvious and also serves as internal accuracy in numerical computation.

Lemma 1. *We have*

$$\boldsymbol{\pi}[\boldsymbol{e}(2S-s+1)\otimes I_{mn}] = \sum_{j=0}^{2S-s} \pi_j = \mu\boldsymbol{\beta}(-T)^{-1}\otimes\boldsymbol{\eta}. \tag{13}$$

Proof. First verify that by definition the vector π satisfies the following equations:

$$\pi_0[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D-\theta I]+\pi_1(I\otimes D_1)=0, \tag{14}$$

$$\pi_j[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D_0-\theta I]+\pi_{j+1}(I\otimes D_1)=0,\ 1\le j\le s-1, \tag{15}$$

$$\pi_s[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D_0-\theta I]+(\pi_{s+1}+\pi_{S+1}+\pi_{S+S_2+1})(I\otimes D_1)=0, \tag{16}$$

$$\pi_j[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D_0]+\pi_{j+1}(I\otimes D_1)=0,\ s+1\le j\le S-S_1, \tag{17}$$

$$\pi_j[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D_0]+\pi_{j+1}(I\otimes D_1)+\theta_2\pi_{j+S_1+S_2}=0,\ S-S_1+1\le j\le S-S_2, \tag{18}$$

$$\pi_j[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D_0]+\pi_{j+1}(I\otimes D_1)+\theta_2\pi_{j+S_1+S_2}+\theta_1\pi_{j+S_2}=0,\ S-S_2+1\le j\le S-1, \tag{19}$$

$$\pi_S[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D_0]+\theta_2\pi_{S+S_1+S_2}+\theta_1\pi_{S+S_2}=0, \tag{20}$$

$$\pi_j[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D_0-\theta_1 I]+\pi_{j+1}(I\otimes D_1)=0,\ S+1\le j\le S+S_2-s-1, \tag{21}$$

$$\pi_j[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D_0-\theta_1 I]+\pi_{j+1}(I\otimes D_1)+\theta_2\pi_{j-S-S_2+s}=0,\ S+S_2-s\le j\le S+S_2-1, \tag{22}$$

$$\pi_{S+S_2}[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D_0-\theta_1 I]+\theta_2\pi_s=0, \tag{23}$$

$$\pi_j[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D_0-\theta_2 I]+\pi_{j+1}(I\otimes D_1)=0,\ S+S_2+1\le j\le S+S_1+S_2-s-1, \tag{24}$$

$$\pi_j[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D_0-\theta_2 I]+\pi_{j+1}(I\otimes D_1)+\theta_1\pi_{j-S-S_1-S_2+s}=0,\ S+S_1+S_2-s\le j\le 2S-s-1, \tag{25}$$

$$\pi_{2S-s}[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D_0-\theta_2 I]+\theta_1\pi_s=0, \tag{26}$$

and the normalizing equation is

$$\sum_{j=0}^{2S-s}\pi_j\boldsymbol{e}=1. \tag{27}$$

Now adding the equations (14-26) and after some simplifications we get

$$\sum_{j=0}^{2S-s}\pi_j[(T+\boldsymbol{T^0}\boldsymbol{\beta})\oplus D]=0, \tag{28}$$

from which the stated result follows immediately.

3.2. Steady-state analysis

Here, we will perform the steady-state analysis of the model under study. Let $\boldsymbol{x}$ be the steady-state probability vector of Q. That is, $\boldsymbol{x}$ satisfies

$$\boldsymbol{x}Q = 0, \quad \boldsymbol{xe} = 1. \tag{29}$$

We partition this vector as:

$$\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \cdots)$$

where $\boldsymbol{x}_0$ is of dimension $(2S - s + 1)m$ giving the steady-state probability vector that the server is idle with inventory at various levels and the arrival process in one of m phases, and $\boldsymbol{x}_i, i \geq 1$ is of dimension $(2S - s + 1)mn$ giving the steady-state probability vector that the server is busy with inventory at various levels, the arrival process in one of m phases and the service is in one of n phases.

Under the stability condition given in (11) the steady state probability vector x is obtained (see, e.g., Neuts [26]) as follows:

$$\begin{gathered} \boldsymbol{x}_0 B_1 + \boldsymbol{x}_1 B_2 = 0, \\ \boldsymbol{x}_0 B_0 + \boldsymbol{x}_1 [B_2 + RA_2] = 0, \\ \boldsymbol{x}_i = \boldsymbol{x}_1 R^{i-1}, \ i \geq 1, \\ \boldsymbol{x}_0 \boldsymbol{e} + \boldsymbol{x}_1 (I - R)^{-1} \boldsymbol{e} = 1. \end{gathered} \tag{30}$$

where the matrix R is the minimal nonnegative solution to the matrix quadratic equation:

$$R^2 A_2 + RA_1 + A_0 = 0. \tag{31}$$

After we discuss the computation of R we will revisit the computation and results related to the steady-state vector $\boldsymbol{x}$.

3.3. Computation of R

When the dimension of R is of reasonable size, one can use a number of well-known methods such as logarithmic reduction (Latouche and Ramaswami [19]) to compute it. However, when the dimension is prohibitively large, one should employ (block) Gauss-Siedel iteration by exploiting the special structure of the coefficient matrices A_0, A_1, and A_2. We will illustrate this below.

3.3.1. Logarithmic reduction algorithm for R

First, we briefly look at the key steps in the logarithmic reduction method [19].

Step 0: $H \leftarrow (-A_1)^{-1} A_0$, $L \leftarrow (-A_1)^{-1} A_2$, $G = L$, $T = H$.

Step 1: $U = HL + LH$; $M = H^2$; $H \leftarrow (I - U)^{-1} M$; $M \leftarrow L^2$;

$$L \leftarrow (I-U)^{-1}M;\;\; G \leftarrow G+TL;\;\; T \leftarrow TH.$$

Continue Step 1 until $\| e - Ge \|_\infty < \varepsilon$.

Step 2: $R = -A_0(A_1 + A_0G)^{-1}$.

3.3.2. (Block) Gauss-Seidel method for R

When the dimension of R is prohibitively large, one should exploit the special structure of the coefficient matrices, $A_0, A_1,$ and A_2. A good procedure to exploit the special structure is Gauss-Seidel method (Stewart [35]). Here we will outline that exploitation. First note that (due to the fact that $RA_2\boldsymbol{e} = A_0\boldsymbol{e} = (\boldsymbol{e}-\boldsymbol{e}_1)(\boldsymbol{e}\otimes D_1\boldsymbol{e}),$ the first block row of R is zero. That is, $R_{0j} = 0, j = 0,\cdots,2S-s$.

Suppose that $V = (V_{ij}) = R^2$. Note that $R^2A_2 = (V_{ij}(\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}\otimes I))$. Now the equation (31) can be written in terms of matrices of dimension mn. In the following, $1 \le i \le 2S-s$ and the notation δ stands for Kronecker delta. That is, $\delta_{i,j} = 1$ for $i = j$, and $\delta_{i,j} = 0$ for $i \ne j$.

$$R_{i0} = [V_{i0}(\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}\otimes I) + \delta_{i1}(I\otimes D_1)][\theta I - (T\oplus D)]^{-1}, \tag{32}$$

$$R_{ij} = [V_{ij}(\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}\otimes I) + \delta_{i-1,j}(I\otimes D_1)][\theta I - (T\oplus D_0)]^{-1},\; 1\le j\le s, \tag{33}$$

$$R_{ij} = [V_{ij}(\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}\otimes I) + \delta_{i-1,j}(I\otimes D_1)][-(T\oplus D_0)]^{-1},\; s+1\le j\le S_2+s, \tag{34}$$

$$R_{ij} = [V_{ij}(\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}\otimes I) + \delta_{i-1,j}(I\otimes D_1) + \theta_2 R_{i,j+S_1+S_2}][-(T\oplus D_0)]^{-1},\; S_2+s+1\le j\le S_1+s \tag{35}$$

$$R_{ij} = [V_{ij}(\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}\otimes I) + \delta_{i-1,j}(I\otimes D_1) + \theta_1 R_{i,j+S_2} + \theta_2 R_{i,j+S_1+S_2}][-(T\oplus D_0)]^{-1}], \quad S_1+s+1\le j\le S-1, \tag{36}$$

$$R_{iS} = [V_{iS}(\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}\otimes I) + \theta_1 R_{i,S+S_2} + \theta_2 R_{i,2S-s}][-(T\oplus D_0)]^{-1}], \tag{37}$$

$$R_{ij} = [V_{ij}(\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}\otimes I) + \delta_{i-1,j}(I\otimes D_1)][\theta_1 I - (T\oplus D_0)]^{-1}],\; S+1\le j\le S+S_2-s-1, \tag{38}$$

$$V_{ij}(\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}\otimes I) + R_{ij}(T\oplus D_0 - \theta_1 I) + \delta_{i-1,j}(I\otimes D_1) + \theta_2 R_{i,j-S-S_2+s} = 0, \quad S+S_2-s\le j\le S+S_2, \tag{39}$$

$$V_{ij}(\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}\otimes I) + R_{ij}(T\oplus D_0 - \theta_2 I) + \delta_{i-1,j}(I\otimes D_1) = 0, \quad S+S_2+1\le j\le S+S_1+S_2-s-1, \tag{40}$$

$$V_{ij}(\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta}\otimes I) + R_{ij}(T\oplus D_0 - \theta_2 I) + \delta_{i-1,j}(I\otimes D_1) + \theta_1 R_{i,j-S-S_1-S_2+s} = 0, \quad S+S_1+S_2-s\le j\le 2S-s. \tag{41}$$

3.4. Computation of the steady-state vector

Once the matrix R is computed, we can evaluate the steady-state probability vector $\boldsymbol{x}$. Once again, depending on the dimension of the problem under study, we can exploit the structure of the coefficient matrices appearing in (30). We further partition $\boldsymbol{x}_i,\ i \geq 0,$ as follows.

$$\boldsymbol{x}_i = (\boldsymbol{x}_{i,0}, \cdots, \boldsymbol{x}_{i,2S-s}),\ i \geq 0.$$

Note that, for $0 \leq j \leq 2S-s$, $\boldsymbol{x}_{0,j}$ is of dimension m whereas $\boldsymbol{x}_{i,j}$ is of dimension mn.

First, we establish a few lemmas which will be useful as internal accuracy check in numerical computation.

Lemma 2. *We have*

$$\sum_{j=0}^{2S-s} [\boldsymbol{x}_{0,j} + \sum_{i=1}^{\infty} \boldsymbol{x}_{i,j}(\boldsymbol{e} \otimes I)] = \boldsymbol{\eta}, \tag{42}$$

where $\boldsymbol{\eta}$ is as given in (1).

Proof. Suppose we write the steady-state equations $\boldsymbol{x}Q = \boldsymbol{0}$ as

$$\begin{aligned} \boldsymbol{x}_0 B_1 + \boldsymbol{x}_1 B_2 &= \boldsymbol{0}, \\ \boldsymbol{x}_0 B_0 + \boldsymbol{x}_1 A_1 + \boldsymbol{x}_2 A_2 &= \boldsymbol{0}, \\ \boldsymbol{x}_{i-1} A_0 + \boldsymbol{x}_i A_1 + \boldsymbol{x}_{i+1} A_2 &= \boldsymbol{0},\ i \geq 2. \end{aligned} \tag{43}$$

Now post-multiplying the first equation in (43) by $(\boldsymbol{e}(2S-s+1) \otimes I_m)$ and the rest of the equations in (43) by $(\boldsymbol{e}(2S-s+1) \otimes \boldsymbol{e}(n) \otimes I_m)$ and adding the resulting equations, we get

$$\sum_{j=0}^{2S-s} [\boldsymbol{x}_{0,j} + \sum_{i=1}^{\infty} \boldsymbol{x}_{i,j}(\boldsymbol{e} \otimes I)]D = \boldsymbol{0}, \tag{44}$$

from which the stated results when using the normalizing condition as well as the steady-state vector of D.

Note. The result in Lemma 2 confirms that in steady-state the phase of the arrival process obtained in a round-about way should be equal to the one obtained directly.

Lemma 3. *We have*

$$\sum_{i=1}^{\infty} \sum_{j=0}^{2S-s} \boldsymbol{x}_{i,j}(I \otimes \boldsymbol{e}) = \mu(1 - \boldsymbol{x}_0 \boldsymbol{e})\boldsymbol{\beta}(-T)^{-1}. \tag{45}$$

Proof. Now post-multiplying the all but the first equation in (43) by $(\boldsymbol{e}(2S-s+1) \otimes I_n \otimes \boldsymbol{e}(m))$ and adding the resulting equations, we get

$$\sum_{i=1}^{\infty}\sum_{j=0}^{2S-s}\boldsymbol{x}_{i,j}(I\otimes\boldsymbol{e})(T+\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta})=0, \tag{46}$$

from which the stated result follows upon using the definition of the steady-state vector of $(T+\boldsymbol{T}^{\mathbf{0}}\boldsymbol{\beta})$, its uniqueness, and the normalizing condition of $\boldsymbol{x}$.

Suppose that $\boldsymbol{z}$, a vector of dimension $(2S-s+1)m$ partitioned into vectors of dimension m as $\boldsymbol{z}=(\boldsymbol{z}_0,\boldsymbol{z}_1,\cdots,\boldsymbol{z}_{2S-s})$. Note that the k^{th} component of the vector $\boldsymbol{z}_j$, for $0\le j\le 2S-s, 1\le k\le m$, gives the steady-state probability that the inventory level is j and the arrival process is in phase k. The following lemma gives an interesting and intuitively clear expression for z in terms of $\boldsymbol{\pi}$ and thus showing that the probability mass function of the inventory level as well as certain probabilities and mean cycle times associated with the vendors are independent of the service time distributions.

Lemma 4. *We have*

$$\boldsymbol{z}_j=\boldsymbol{\pi}_j(\boldsymbol{e}\otimes I),\ 0\le j\le 2S-s. \tag{47}$$

Proof. First verify that

$$B_1+B_0(\boldsymbol{e}\otimes I)=B_2+(A_0+A_1)(\boldsymbol{e}\otimes I)=(A_0+A_1+A_2)(\boldsymbol{e}\otimes I). \tag{48}$$

By definition of z, it is easy to see that

$$\boldsymbol{z}_j=\boldsymbol{x}_{0,j}+\sum_{i=1}^{\infty}\boldsymbol{x}_{i,j}(\boldsymbol{e}\otimes I),\ 0\le j\le 2S-s. \tag{49}$$

Now post-multiplying the second and third equations in (43) by $(\boldsymbol{e}\otimes I)$ and adding the resulting equations with the first equation in (43), and with the help of (48) and the definition of $\boldsymbol{z}$, we get the following equations.

$$\boldsymbol{z}_0(D-\theta I)+\boldsymbol{z}_1D_1=\mathbf{0}, \tag{50}$$

$$\boldsymbol{z}_j(D_0-\theta I)+\boldsymbol{z}_{j+1}D_1=0,\ 1\le j\le s-1, \tag{51}$$

$$\boldsymbol{z}_s(D_0-\theta I)+(\boldsymbol{z}_{s+1}+\boldsymbol{z}_{S+1}+\boldsymbol{z}_{S+S_2+1})D_1=0, \tag{52}$$

$$\boldsymbol{z}_jD_0+\boldsymbol{z}_{j+1}D_1=0,\ s+1\le j\le S-S_1, \tag{53}$$

$$\boldsymbol{z}_jD_0+\boldsymbol{z}_{j+1}D_1+\theta_2\boldsymbol{z}_{j+S_1+S_2}=0,\ S-S_1+1\le j\le S-S_2, \tag{54}$$

$$\boldsymbol{z}_jD_0+\boldsymbol{z}_{j+1}D_1+\theta_2\boldsymbol{z}_{j+S_1+S_2}+\theta_1\boldsymbol{z}_{j+S_2}=0,\ S-S_2+1\le j\le S-1, \tag{55}$$

$$\boldsymbol{z}_SD_0+\theta_2\boldsymbol{z}_{S+S_1+S_2}+\theta_1\boldsymbol{z}_{S+S_2}=0, \tag{56}$$

$$\boldsymbol{z}_j(D_0-\theta_1I)+\boldsymbol{z}_{j+1}D_1=0,\ S+1\le j\le S+S_2-s-1, \tag{57}$$

$$z_j(D_0-\theta_1 I)+z_{j+1}D_1+\theta_2 z_{j-S-S_2+s}=0, S+S_2-s\le j\le S+S_2-1, \quad (58)$$

$$z_{S+S_2}(D_0-\theta_1 I)+\theta_2 z_s=0, \quad (59)$$

$$z_j(D_0-\theta_2 I)+z_{j+1}D_1=0,\ S+S_2+1\le j\le S+S_1+S_2-s-1, \quad (60)$$

$$z_j(D_0-\theta_2 I)+z_{j+1}D_1+\theta_1 z_{j-S-S_1-S_2+s}=0,\ S+S_1+S_2-s\le j\le 2S-s-1, \quad (61)$$

$$z_{2S-s}(D_0-\theta_2 I)+\theta_1 z_s=0. \quad (62)$$

The above equations have identical structure to those listed in (14-26) in that we get (50-62), respectively, by post-multiplying each one of (14-26) by $e\otimes I$. This fact along with the uniqueness of π and $ze=1$ yield the stated result.

Lemma 5. *The probability, P_{loss}, that an arriving customer is lost due to zero inventory is given by*

$$P_{loss}=\frac{1}{\lambda}\pi_0(e\otimes D_1 e). \quad (63)$$

Proof. Note that an arriving customer is lost if and only at that time the inventory level is zero. Thus, by definition, we have

$$P_{loss}=\frac{1}{\lambda}[x_{0,0}D_1 e+\sum_{i=1}^{\infty}x_{i,0}(e\otimes D_1 e)], \quad (64)$$

from which, after applying Lemma 4 (by taking $j=0$), the stated result follows immediately.

Note. The effective arrival rate is thus given by $\lambda-\pi_0(e\otimes D_1 e)$. Hence, the stability condition given in (11) is intuitively clear.

3.5. System performance measures

In this section we list a number of key system performance measures along with their expressions to bring out the qualitative aspects of the model under study. In the sequel, we define the probability vector, $\xi=(\xi_0,\cdots,\xi_{2S-s})$ of dimension $2S-s+1$ as

$$\xi=x_0(I\otimes e(m))+x_1(I-R)^{-1}(I\otimes e(mn))=z(I\otimes e(m)).$$

Thus, ξ gives the probability mass function of the inventory level.

1. *Probability that the system is idle.* The probability, P_{idle}, that the system is idle (i.e. there are no customers in the system) at an arbitrary time is given by

$$P_{idle}=x_0 e=1-\rho.$$

Note that the last equality follows from Lemma 3 and the fact that in steady-state the input and output rates should be equal.

2. *Probability that the server is busy.* The probability P_{busy}, that the server is busy is given by

$$P_{busy} = 1 - \boldsymbol{x}_0 \boldsymbol{e} = \rho.$$

3. *Mean number of customers in the system.* The mean, μ_{NS}, number of customers in the system is given by

$$\mu_{NS} = \sum_{i=1}^{\infty} i \boldsymbol{x}_i \boldsymbol{e} = \boldsymbol{x}_1 (I - R)^{-2} \boldsymbol{e}.$$

4. *Mean inventory level.* The mean, μ_{IL}, number of inventory in the system is given by

$$\mu_{IL} = \sum_{i=1}^{S} i\, \xi_i + \sum_{i=s+1}^{s+S_2} i\, \xi_{S-s+i} + \sum_{i=s+1}^{s+S_1} i\, \xi_{S+S_2-s+i}.$$

5. The probability, P_{loss}, that an arriving customer is lost due to zero inventory is as given in (47).

6. The quantities, $f_{pending}^{(12)}, f_{pending}^{(1)}, f_{pending}^{(2)}$, respectively, representing the fraction of time both vendors, Vendor 1, and Vendor 2, have pending replenishment at an arbitrary time are computed as follows.

$$f_{pending}^{(12)} = \sum_{i=0}^{s} \xi_i,$$

$$f_{pending}^{(1)} = \sum_{i=0}^{s} \xi_i + \sum_{i=s+1}^{s+S_2} \xi_{S-s+i},$$

$$f_{pending}^{(2)} = \sum_{i=0}^{s} \xi_i + \sum_{i=s+1}^{s+S_1} \xi_{S+S_2-s+i}.$$

7. *Mean cycle time.* The mean cycle time, $\mu_{CT}^{(i)}$, of Vendor $i, i = 1, 2$, defined as the mean time between two successive replenishment by Vendor i, is obtained as

$$\mu_{CT}^{(i)} = [\theta_i f_{pending}^{(i)}]^{-1}, \; i = 1, 2.$$

4. Illustrative Numerical Examples

In this section we will discuss a few representative examples to bring out the qualitative nature of the model under study. Towards this end, we consider several *MAPs* covering renewal and correlated inter-demand times, and three $PH-$distributions for services.

Arrival processes: For arrival process we consider several $MAPs$ as listed below.

1. **Erlang** (*ERA*):

$$D_0 = \begin{bmatrix} -2 & 2 \\ 0 & -2 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix}.$$

2. **Exponential** (*EXP*):

$$D_0 = [-1], \quad D_1 = [1].$$

3. **Hyperexponential** (*HEA*):

$$D_0 = \begin{bmatrix} -8.2 & 0 & 0 \\ 0 & -0.82 & 0 \\ 0 & 0 & -0.082 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 5.74 & 2.05 & 0.41 \\ 0.574 & 0.205 & 0.041 \\ 0.0574 & 0.0205 & 0.0041 \end{bmatrix}.$$

4. **Negatively Correlated 1** (NC_1):

$$D_0 = \begin{bmatrix} -1.25 & 1.25 & 0 \\ 0 & -1.25 & 0 \\ 0 & 0 & -2.5 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0.0125 & 0 & 1.2375 \\ 2.475 & 0 & 0.025 \end{bmatrix}.$$

5. **Poisitively Correlated 1** (PC_1):

$$D_0 = \begin{bmatrix} -1.25 & 1.25 & 0 \\ 0 & -1.25 & 0 \\ 0 & 0 & -2.5 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0 \\ 1.2375 & 0 & 0.0125 \\ 0.025 & 0 & 2.475 \end{bmatrix}.$$

6. **Negatively Correlated 2** (NC_2):

$$D_0 = \begin{bmatrix} -1.75 & 1.75 & 0 & 0 \\ 0 & -1.75 & 1.75 & 0 \\ 0 & 0 & -1.75 & 0 \\ 0 & 0 & 0 & -3.5 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.0175 & 0 & 0 & 1.7325 \\ 3.465 & 0 & 0 & 0.035 \end{bmatrix}.$$

7. **Positively Correlated 2 (**PC_2**):**

$$D_0 = \begin{bmatrix} -1.75 & 1.75 & 0 & 0 \\ 0 & -1.75 & 1.75 & 0 \\ 0 & 0 & -1.75 & 0 \\ 0 & 0 & 0 & -3.5 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1.7325 & 0 & 0 & 0.0175 \\ 0.035 & 0 & 0 & 3.465 \end{bmatrix}.$$

8. Negatively Correlated 3 (NC_3):

$$D_0 = \begin{bmatrix} -2.25 & 2.25 & 0 & 0 & 0 \\ 0 & -2.25 & 2.25 & 0 & 0 \\ 0 & 0 & -2.25 & 2.25 & 0 \\ 0 & 0 & 0 & -2.25 & 0 \\ 0 & 0 & 0 & 0 & -4.5 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.0225 & 0 & 0 & 0 & 2.2275 \\ 4.455 & 0 & 0 & 0 & 0.045 \end{bmatrix}.$$

9. Positively Correlated 3 (PC_3):

$$D_0 = \begin{bmatrix} -2.25 & 2.25 & 0 & 0 & 0 \\ 0 & -2.25 & 2.25 & 0 & 0 \\ 0 & 0 & -2.25 & 2.25 & 0 \\ 0 & 0 & 0 & -2.25 & 0 \\ 0 & 0 & 0 & 0 & -4.5 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 2.2275 & 0 & 0 & 0 & 0.0225 \\ 0.045 & 0 & 0 & 0 & 4.455 \end{bmatrix}.$$

The above arrival processes will be normalized in order to have the value for λ. Note that these are qualitatively different. The following table gives the ratio of the standard deviations (to Erlang (ERA)) and the 1-lag correlation coefficients (1-lag-cc) of the inter-arrival times.

Table 1: Arrival processes statistics

TaP	ERA	EXA	HEA	NC_1	PC_1	NC_2	PC_2	NC_3	PC_3
$Ratio\ SD$	1	1.414	3.175	1.470	1.470	1.443	1.443	1.432	1.432
$1-lag-cc$	0	0	0	-0.327	0.327	-0.480	0.480	-0.580	0.579

Service times: We look at the following three service times. Again, these distributions will be normalized to a desired value for μ, and also these are qualitatively different.

1. **Erlang** (*ERS*)**:**

$$\beta = (1,0), \quad S = \begin{bmatrix} -2 & 2 \\ 0 & -2 \end{bmatrix}.$$

2. **Exponential** (*EXS*)**:**

$$\beta = [1], \quad S = [-1].$$

3. **Hyperexponential** (*HES*)**:**

$$\beta = [0.9, 0.1], \quad S = \begin{bmatrix} -10 & 0 \\ 0 & -1 \end{bmatrix}.$$

Example 1. In this example, we will explore the behavior of some key measures related to inventory level, replenishment, and cycle times, as we vary s and S_1 by fixing $S = 200, \lambda = 1, \theta_1 = \theta_2 = 0.1$ and considering the nine *MAPs* listed earlier. Note that for this example the service time does not play any role. In Figures 2-9, we display the measures, $\lambda_e = \lambda(1 - P(loss)), P(loss), f_{pending}^{(12)}, f_{pending}^{(1)}, f_{pending}^{(2)}, \mu_{CT_1}^{(1)},$ and $\mu_{CT_2}^{(2)}$. From these figures, we notice the following observations.

- The measure, λ_e, the effective arrival rate, approaches $\lambda = 1$ as s is increased for all but positively correlated arrival processes. For positively correlated arrival processes it appears that one needs a larger values for s and S.
- Keeping the value of the service rate constant, it can be observed that at low values of s, varying the value of S_1 (the amount of ordered sent to Vendor 1) can cause the probability of lost demand to increase by almost 10%. However at higher values of s, this change is non-existent. Not only that but the higher the value of s, the lower the probability of lost demand which makes sense intuitively as larger minimum inventory threshold would mean that the likelihood an order can be fulfilled is high. However for reducing the inventory storage costs, a lower s level is ideal. In this case an s value of 25 would be most suitable as that keeps the probability of a lost demand below 0.05 for all values of S_1.
- The mean level of inventory in the system linearly increases as s is increased. As S_1 is increased keeping s constant, the mean level of inventory increases at low values of s but is much more stable at high levels of s. Again a value of s close to 25 in this case would be the most suitable value to keep inventory costs low while meeting demand with high probability.
- The mean cycle time for Vendor 1 is significantly impacted by s but also affected by S_1 as a low s and high S_1 indicates a very high cycle time. A high s with any value of S_1 has a lower cycle time primarily because the net order size is reduced hence the cycle time to fulfill the order goes down.
- This effect is even more pronounced using data for Vendor 2 who has a very high cycle time of S_1. This is primarily because we are taking a much larger set of values for Vendor 2 compared to Vendor 1 as Vendor 2 delivers the leftover quantity after the main order is placed with Vendor 1. The proportion of orders placed between two vendors can have an impact up to 50 units on the cycle time which could cause delayed deliveries. In an ideal situation, the proportion should be kept close to the 50/50 mark; however, there are various reasons this is not possible including vendor capacity, location and price points.

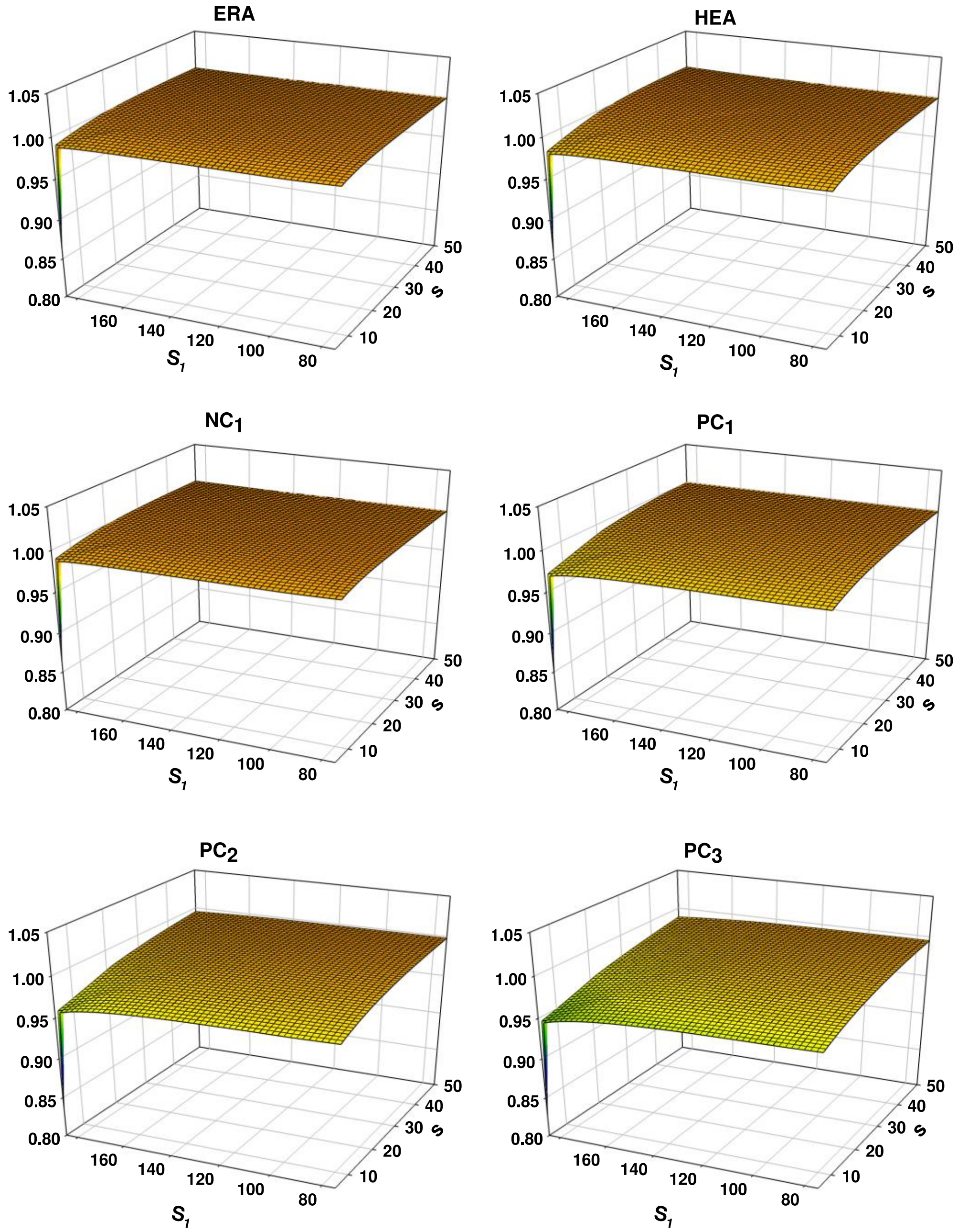

Figure 2. Effective arrival rate under different scenarios.

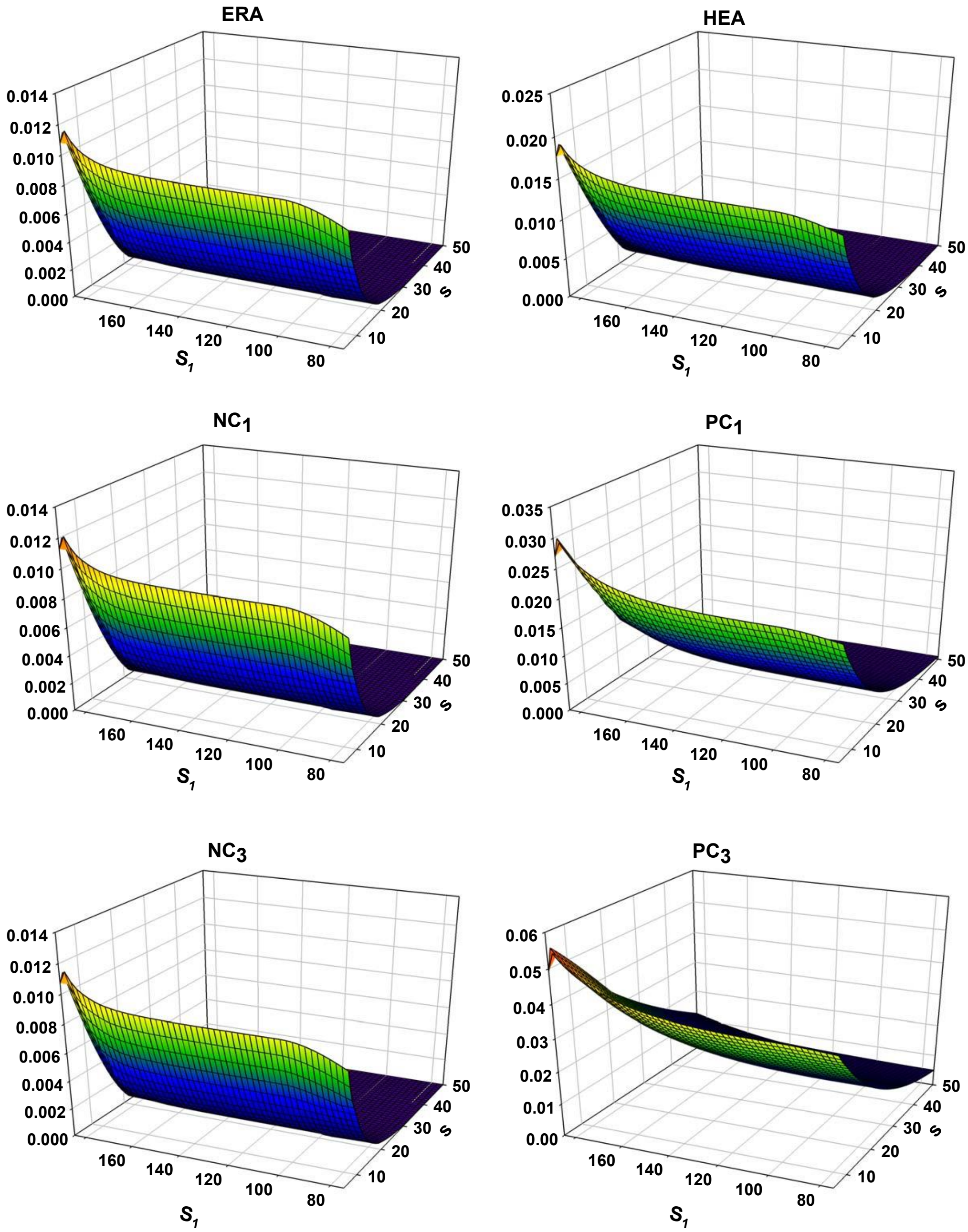

Figure 3. P(loss) under different scenarios.

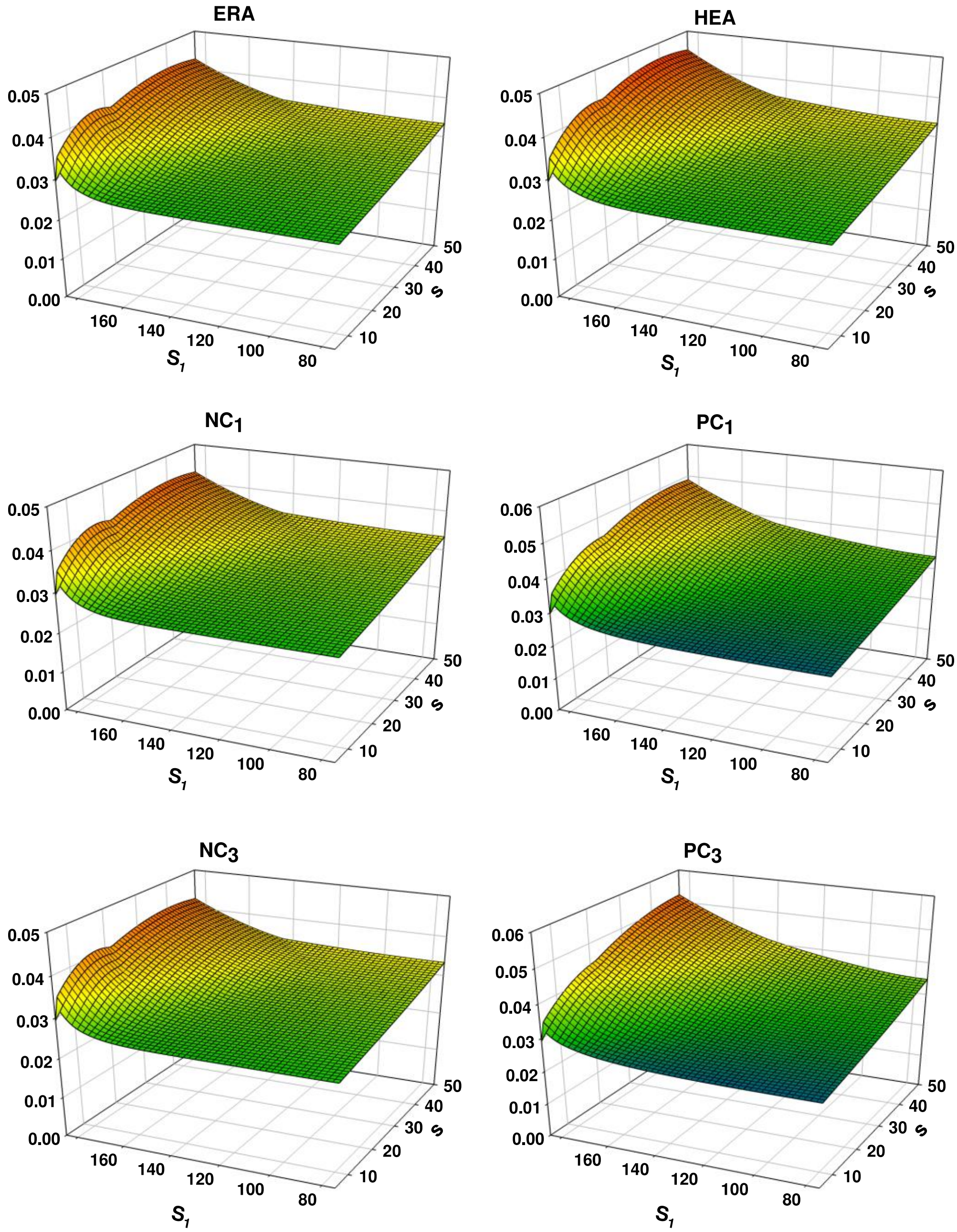

Figure 4. Fraction of time both vendors have pending replenishment under different scenarios.

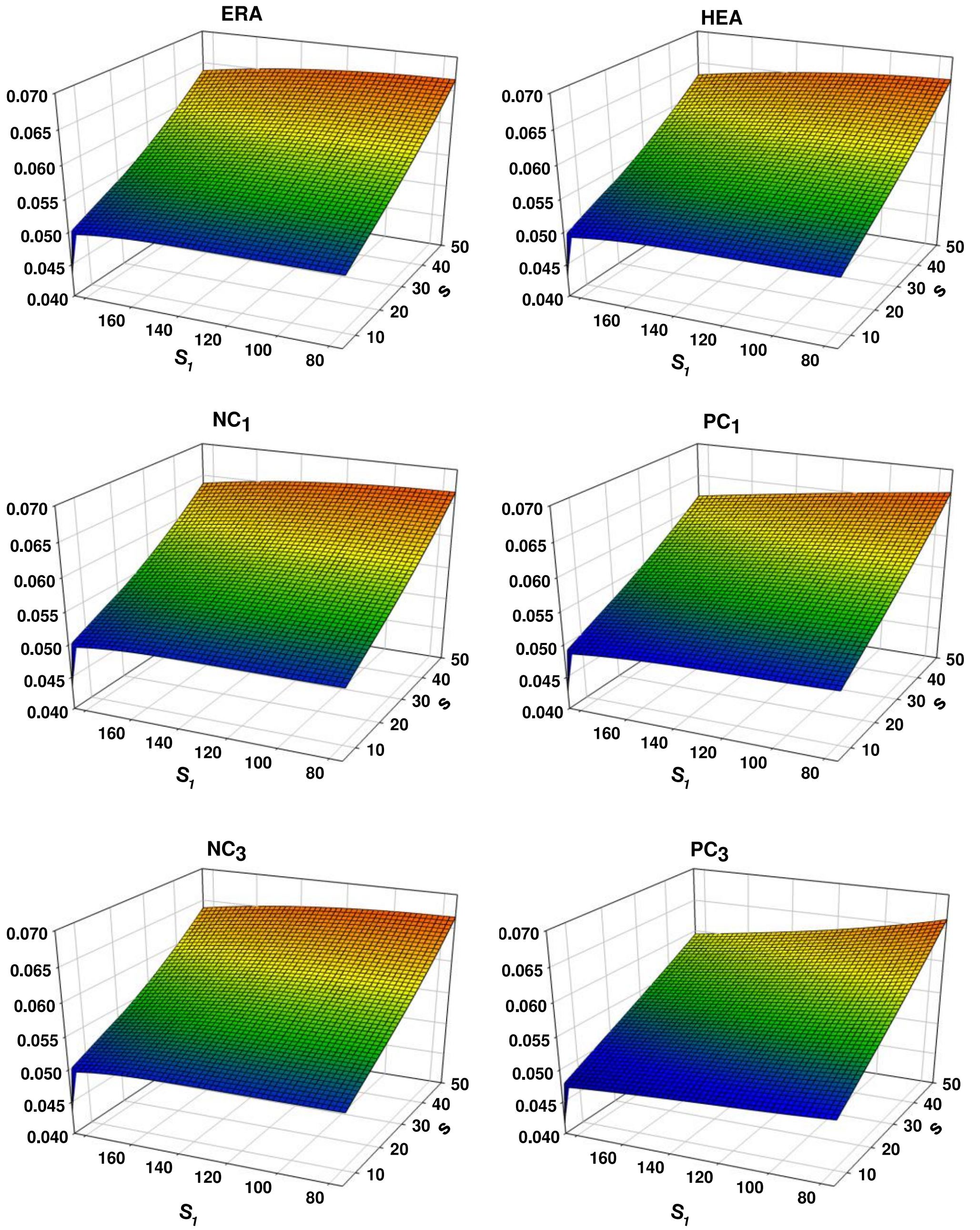

Figure 5. Fraction of time only Vendor 1 has pending replenishment under different scenarios.

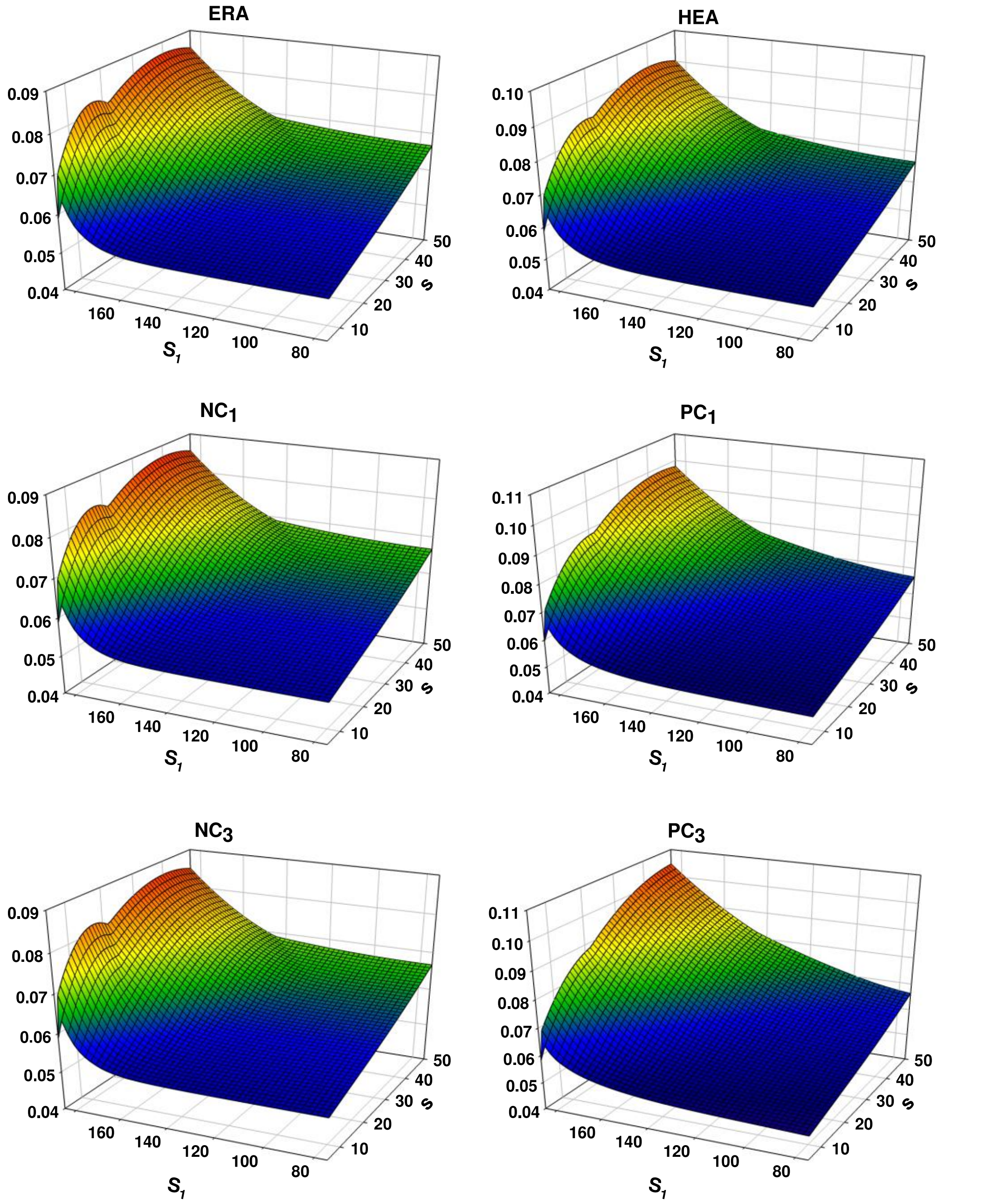

Figure 6. Fraction of time only Vendor 2 has pending replenishment under different scenarios.

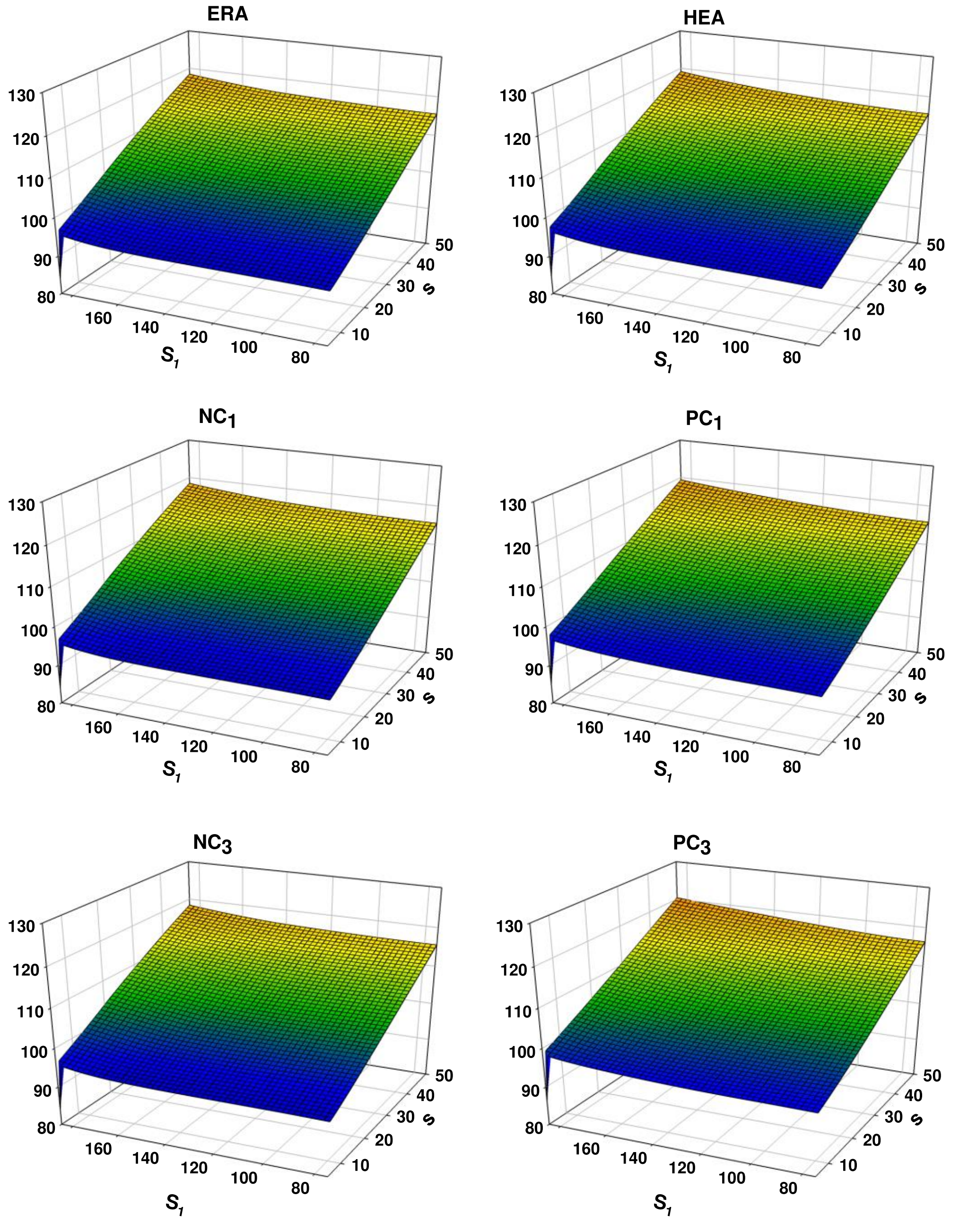

Figure 7. Mean inventory level under different scenarios.

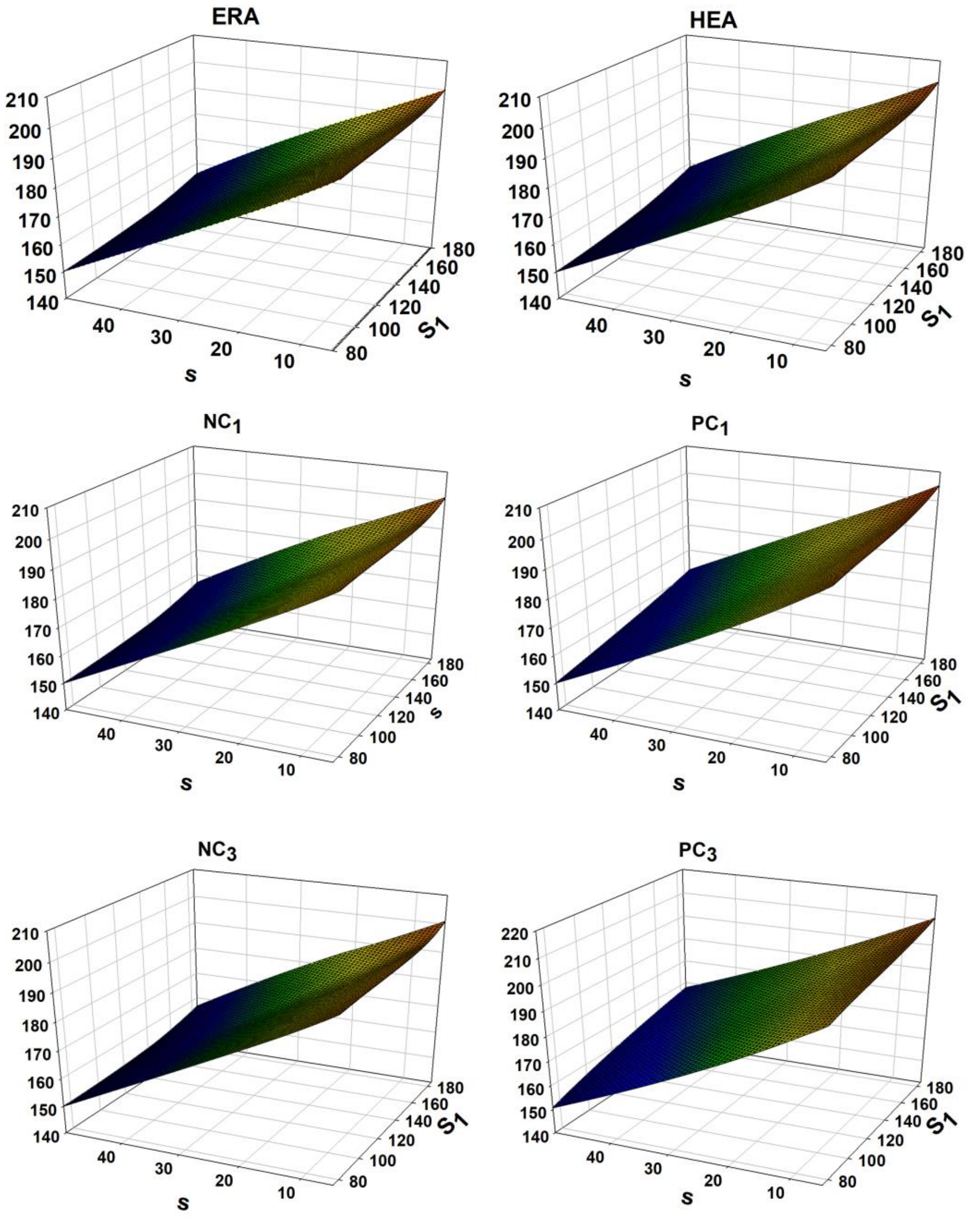

Figure 8. Mean cycle time of Vendor 1 under different scenarios.

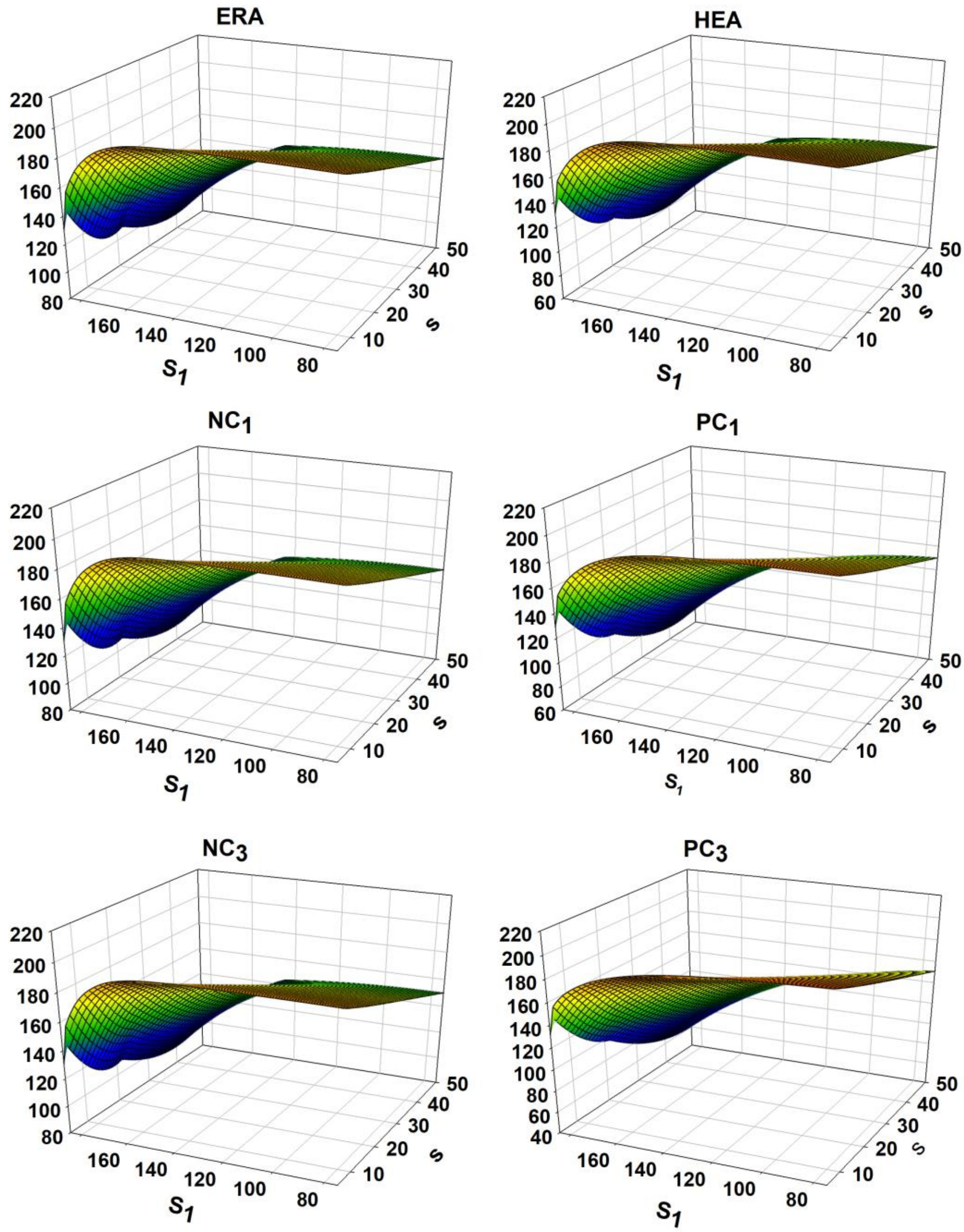

Figure 9. Mean cycle time of Vendor 2 under different scenarios.

Example 2. In this example, we compare one vendor and two-vendor cases. Towards this end, we will vary S_1 by fixing $S = 50, s = 5, \lambda = 1, \theta_1 = \theta_2 = 0.1, \rho = 0.95$ and consider the nine *MAPs* listed earlier for the arrivals of the demands and the three $PH -$distributions for the services. We vary S_1 and also observe that μ will vary as we fix the traffic intensity. Note that one vendor case ($S_2 = 0$) is studied extensively in the literature. In order to do a proper comparison of these models, we will look at the measures common to one vendor model, say, Model 1, and two-vendor model, say, Model 2. We look at the ratios computed as the measure of model 2 over the corresponding measure of model 1. For example, the ratio for P_{loss}, is obtained as $P_{loss}^{Model\,2} / P_{loss}^{Model\,1}$. In Figure 10, we display the ratios corresponding to the service rate and the measures, $P_{loss}, f_{pending}^{(1)}, \mu_{IL}$, and $\mu_{CT_1}^{(1)}$, and the measure μ_{NS} is displayed in Figure 11 under different scenarios. From these figures, we note the following.

1. Under all scenarios we notice that the ratio for P_{loss} is significantly less than 1 indicating that Model 2 has a much smaller loss probability as compared to that of Model 1. This shows that the two-vendor model outperforms that of one vendor case. This is a very notable observation since in a practical supply chain model, the goal is to minimize the probability that an incoming demand is lost as that leads to lost revenue from current customers.

2. Again with respect to the ratio for the P_{loss}, we see that a higher variability in the arrival process (see, e.g., *HEA* arrivals) yields a higher ratio. Also, we notice that while the level of negative correlation appears to not have an impact on this ratio, we see a different trend in the case of positively correlated arrivals. As the (positive) correlation increases, the ratio also increases. This is the case for all S_1. This is a very notable and significant observation since in practice the inter-arrival times of successive demands are correlated and hence one cannot ignore this, especially, when the correlation is positive.

3. The ratio for the measure, $f_{pending}^{(1)}$, decreases as S_1 is increased. Also, this ratio is greater than 1 and decreases to 1 as S_1 increases. This is the case for all scenarios indicating that Model 2 has a higher probability of Vendor 1 has a replenishment pending. This can be intuitively explained as follows. The size of replenishment for Vendor 1 in the two-vendor model is S_1 which is less than $S - s$. This creates a larger rate of replenishment for the two-vendor case as compared to one vendor case. This can also be seen in the figure for the ratio for the mean cycle time for Vendor 1. Since the mean cycle time and the rate of replenishment are inverse to each other, Model 1 having a larger mean cycle time indicates that Model 2 will have a higher rate of replenishment compared to Model 1. Also we notice this ratio

to be somewhat insensitive to the type of arrival process.

4. We see that the mean inventory level for Model 2 is less than that of Model 1. Again, this is an interesting and significant observation as the holding cost for the two-vendor case will be less compared to the corresponding one vendor model.

5. With respect to the ratio for the μ_{IL}, we see that a higher variability in the arrival process (see, e.g., HEA arrivals) yields a higher ratio. Also, we notice that while the level of negative correlation appears to not have an impact on this ratio, we see a different trend in the case of positively correlated arrivals. As the (positive) correlation increases, the ratio also increases. This is the case for all S_1. Once again this is a very notable and significant observation since in practice the inter-arrival times of successive demands are correlated and hence one cannot ignore this, especially, when the correlation is positive.

6. Now looking at the ratio for the mean number in system, μ_{NS}, we notice a very interesting trend. While for the renewal and negatively correlated arrivals, the ratio is less than 1, implying that one vendor model has a larger mean in the system compared to that of the two-vendor case, we see exactly the opposite trend for the positively correlated arrivals.

As Example 2 points out a two-vendor model has the advantage over the corresponding one vendor model in that probability of loss of demands is reduced while at the same time the inventory holding cost will also be smaller. Further, from a practical point of view, having two vendors to replenish inventory will help the management to have a back-up type replenishment process in case of unforeseen circumstances due to nature or otherwise.

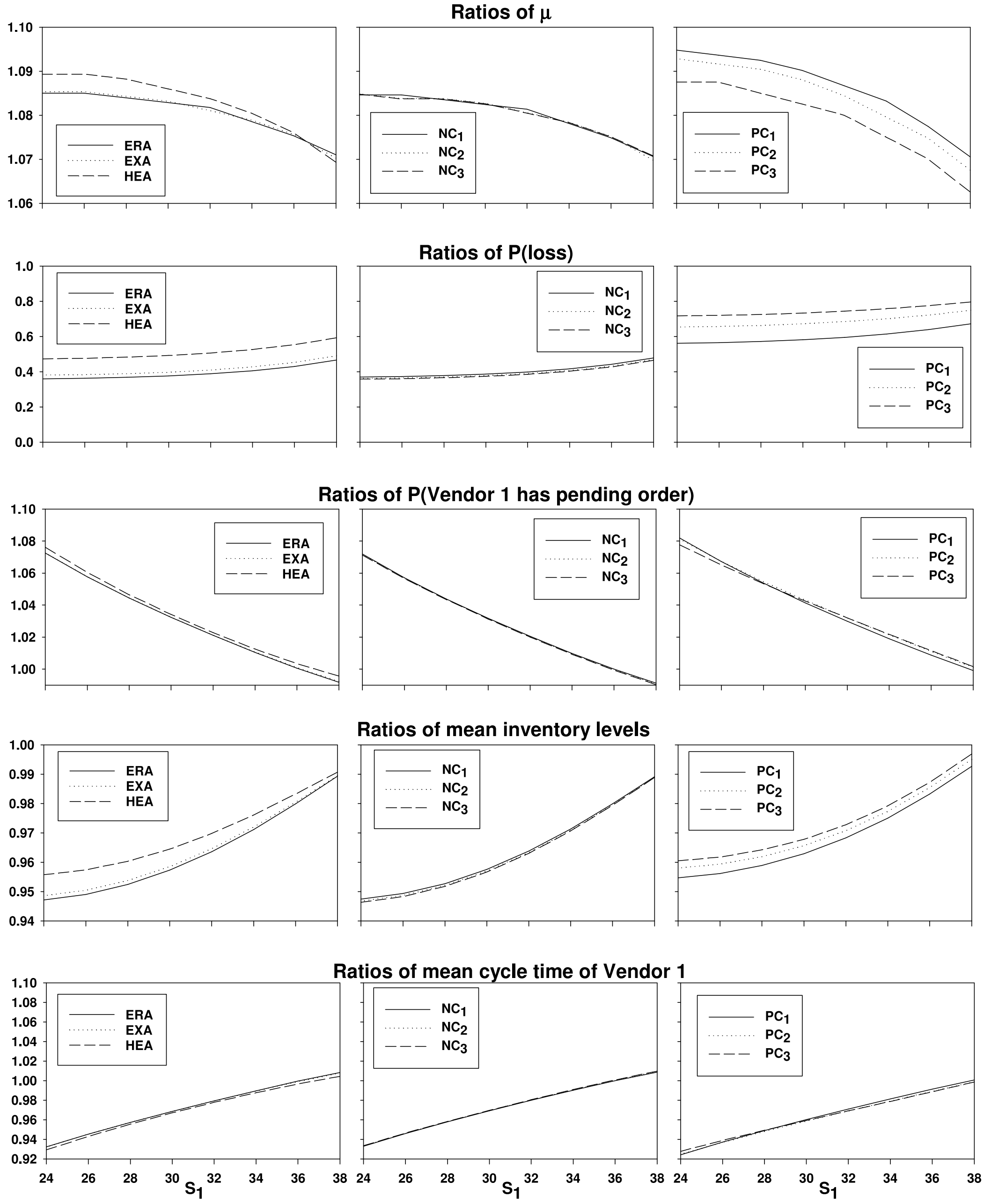

Figure 10. Various ratios of selected measures under different scenarios.

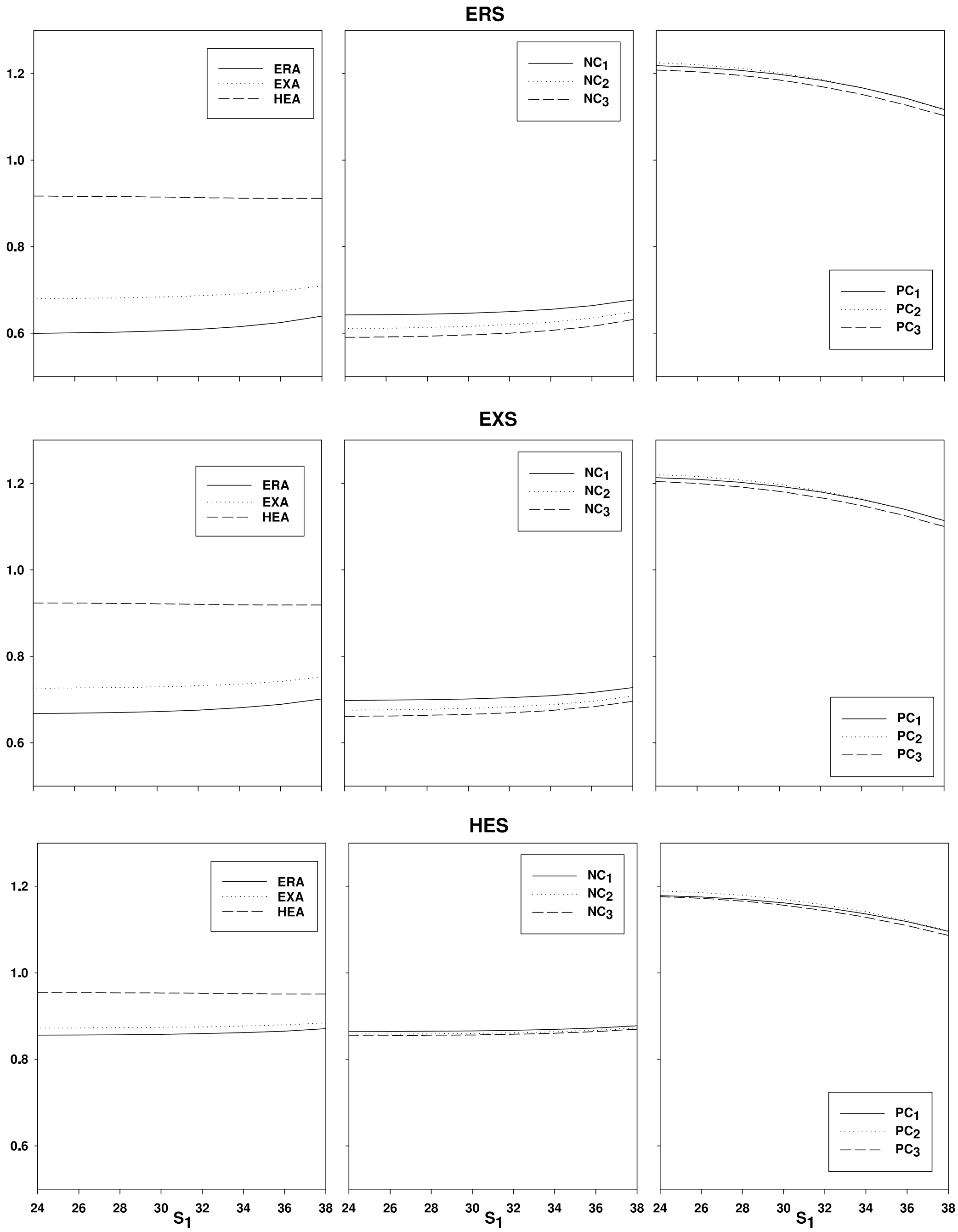

Figure 11. Ratios of the mean number in the system under different scenarios.

5. Concluding Remarks

In this paper we introduced a two-vendor queueing-inventory systems such that replenishment occurs using two vendors as opposed to the classical approach of using only one vendor. Through a versatile arrival process to model the demands, we illustrated the impact of correlation, especially positively correlated ones, in some system performance measures. We also showed the significant advantage of using a two-vendor system as compared to the corresponding one vendor one. The model studied in this paper can be extended in a number of ways. Some of these are: (a) one can look at more than two vendors for replenishment; (b) the assumption of exponential lead times can be relaxed to use phase type distribution. In this case the analysis will be similar except that the dimension of the problem increases significantly; (c) One of the key assumptions in our model is that an incoming demand is lost if the inventory level is zero; however it would be worthwhile to have a buffer (of finite capacity) where an incoming demand that is not immediately met is stored for a finite amount of time before it is considered lost. From a practical point of view this is feasible as there are certain critical items where the orders are fulfilled in such a way. This would introduce further complexity to the model but would help mimic the real-world supply chains further; (d) while we studied our model in steady-state, it would be of interest to study the transient analysis so as to model the seasonal fluctuations (holiday seasons such as Christmas and thanksgiving in the United States when the demand suddenly spikes and a larger number of orders are expected to be fulfilled in a shorter amount of time), which are seen in some real-life applications and which can have a significant impact in meeting the incoming demands. Further research needs to be conducted on the best approaches to incorporate the time element and measure its impact on the model; (e) considering the fact that the incoming material has to go through manufacturing processes that have their own cycle times, it would be a good approach to assess the impact of the single vendor and multi-vendor case on manufacturing cycle times to gain a deeper understanding of the benefit of a multi-vendor strategy for manufacturing operational efficiency and scheduling. As before, this sort of an analysis is extremely complex and further research needs to be conducted in this area; and finally (f) we can allow the customer's demands to be random with a finite support. These extensions are under investigation and the results will be reported elsewhere.

Acknowledgements

The authors are grateful to the reviewers and the editor for their constructive comments that improved the presentation of the paper.

References

[1] Artalejo, J. R., Gomez-Correl, A., & He, Q. M. (2010). Markovian arrivals in stochastic modelling: a survey and some new results. *SORT*, 34, 101-144.

[2] Benny, B., Chakravarthy, S. R., & Krishnamoorthy, A. (2018). Queueing-inventory System with two commodities. *Journal of the Indian Society for Probability and Statistics*, 19, 437-454.

[3] Berman, O., Kaplan, E.H., & Shimshak, D. G. (1993). Deterministic approximations for inventory management at service facilities. *IIE Transactions*, 25, 98-104.

[4] Bladt, M., & Nielsen, B. F. (2017). Matrix-Exponential Distributions in Applied Probability. Springer LLC.

[5] Bradley, E. L. (1969). Queues with balking and their application to an inventory problem. In *Technical Report*, DTIC Document.

[6] Buchholz, P., Kriege, J., & Felko, I. (2014). Input Modeling with Phase-Type Distributions and Markov Models - Theory and Applications. Springer.

[7] Chakravarthy, S. R. (2001). The batch Markovian arrival process: A review and future work. *Advances in Probability Theory and Stochastic Processes*. Eds., A. Krishnamoorthy et al. Notable Publications Inc., New Jersey, 21-39.

[8] Chakravarthy, S. R. (2010). Markovian Arrival Processes. Wiley Encyclopedia of Operations Research and Management Science. New York: Wiley.

[9] Chakravarthy, S. R. (2015). Matrix-Analytic Queueing Models, Chapter 8 in An Introduction to Queueing Theory, U. Narayan Bhat, Second Edition, Birkhauser, Springer Science + Business Media, New York, 2015.

[10] Chakravarthy, S. R., Maity,A., & Gupta, U. C. (2017). Modeling and analysis of bulk service queues with an inventory under "(s,S)" policy. *Annals of Operations Research*, 258, 263-283. [Published Online: 31 October 2015; DOI:10.1007/s10479-015-2041-z]

[11] Choi K. H., & Yoon B. K. (2016). A survey on the queueing inventory systems with phase-type service distributions, QTNA '16, December 13-15, 2016, Wellington, New Zealand (2016). DOI: http://dx.doi.org/10.1145/3016032.3016033.

[12] Christopher, M., & Towill, D. (2000). Supply chain migration from lean and function to agile and customized. *Supply Chain Management: An International Journal*, 5, 206-213.

[13] Cranfield Management School. (2002). Supply Chain Vulnerability. New York, NY: Cranfield University Press.

[14] Graham, A. (1981). Kronecker Products and Matrix Calculus with Applications. Ellis Horwood, Chichester, UK.

[15] He, Qi-Ming. (2014). Fundamentals of Matrix-Analytic Methods. Springer, New York.

[16] Karthikeyan, K., & Sudhesh, R. (2016). Recent review article on queueing inventory systems. *Research Journal of Pharmacy and Technology*. 9, 1451-1461. DOI:10.5958/0974-360X.2016.00421.2

[17] Krishnamoorthy A., Lakshmy B., & Manikandan R. (2011). A survey on inventory models with positive service time, OPSEARCH, 48, 153-169.

[18] Krishnamoorthy A., Manikandan R., & Lakshmy B. (2015). A revisit to queueing-inventory system with positive service time, *Annals of Operations Research*, 233, 221-236.

[19] Latouche G., & Ramaswami V. (1999). Introduction to Matrix Analytic Methods in Stochastic Modeling. SIAM.

[20] Latour, A. (2001). Trial by fire: a blaze in albuquerque sets off major crisis for cell-phone giants. The Wall Street Journal.

[21] Lucantoni, D., Meier-Hellstern K. S. & Neuts, M. F.(1990). A single server queue with server vacations and a class of nonrenewal arrival processes, *Advance Applied Probability*, 22, 676-705.

[22] Marcus, M., & Minc, H. (1964). A Survey of Matrix Theory and Matrix Inequalities. Allyn and Bacon, Boston, MA.

[23] Nahmias, S. (2009). Production and Operations Analysis, Sixth Edition. McGraw Hill /Irwin, Burr Ridge, Illinois, USA.

[24] Neuts, M. F. (1975). Probability distributions of phase type. In *Liber Amicorum Prof. Emeritus H. Florin,* Department of Mathematics, University of Louvain, 173-206.

[25] Neuts, M. F. (1979). A versatile Markovian point process. *Journal of Applied Probability*, 16, 764-779.

[26] Neuts, M. F. (1981). Matrix-Geometric Solutions in Stochastic Models:An Algorithmic Approach. The Johns Hopkins University Press, Baltimore, MD. [1994 version is Dover Edition].

[27] Neuts, M. F. (1989). Structured Stochastic Matrices of M/G/1 Type and Their Applications. CRC Press.

[28] Neuts, M. F. (1992). Models based on the Markovian arrival process, *IEICE Transactions on Communications*, E75B, 1255-1265.

[29] Reitman, V.(1997). To the rescue: Toyota's fast rebound after fire at supplier shows why it is tough. *The Wal Street Journal*.

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

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

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

[33] Sheffi, Y. (2005). The Resilient Enterprise. Cambridge, MA, MIT Press.

[34] Steeb, W.-H., & Hardy, Y. (2011). Matrix Calculus and Kronecker Product. World Scientific Publishing, Singapore.

[34] Stewart W. J. (1994). Introduction to the Numerical Solution of Markov Chains. Princeton, New Jersey, Princeton University Press.

[36] Tang, C. (2006). Perspectives in supply chain risk management. *International Journal of Production Economics*, 103, 451- 88.

[37] Wong, C., McFarlane, D., & Zaharudin, A. A. (2002). The intelligent product driven supply chain. *Proceedings of IEEE Internation Conference on Systems, Man and Cybernetics*, Tunisia.

[38] Yu, H., Zeng, A. Z., & Zhao, L. (2009). Single or dual sourcing: decision-making in the presence of supply chain disruption risks. *Omega*, 37, 788-800.

[39] Yue, D., & Qin, Y. (2019). A production inventory system with service time and production vacations. *Journal of Systems Science and Systems Engineering*, 28, 168-180.