Operational Research (2025) 25:90
https://doi.org/10.1007/s12351-025-00970-w

ORIGINAL PAPER

Performance analysis and optimization of state-dependent replenishment policy in queuing-inventory system

Agassi Melikov[1] · **Serife Ozkar**[2]

Received: 6 May 2024 / Revised: 22 July 2025 / Accepted: 31 July 2025 /
Published online: 2 September 2025

Abstract

This study proposes a new queue-dependent (s, S)-type replenishment policy in the queueing-inventory systems. If the inventory level is greater than s, no restocking order is sent; otherwise, the replenishment is performed to reach the maximum value S, regardless of the stock level at the moment. Orders are replenished in two ways: the regular order and the urgent order. Lead times of the two-type orders are exponentially distributed with different parameters. The urgent orders require a shorter delivery time than the regular orders. The queue-dependent replenishment policy is defined as follows: when the inventory level drops to s, if the number of the customers is less than predefined threshold value r, the regular order is made; if the number is more or equal r, the urgent order is sent. When the inventory level drops to zero, one customer becomes impatient, regardless of the customer's number in the queue. Arrival of customers is according to a Markovian arrival process and the service times are adapted by a phase-type distribution. The mathematical model of the system is developed using a continuous-time Markov chain with an infinite state space. Stability condition and then the steady-state distribution are derived by using the matrix-geometric method. The influences of the parameters on the performance measures are discussed with numerical examples. An optimization problem is solved, where the criterion is the expected total cost, and the controlled parameters are the reorder point s and the threshold parameter r.

Keywords Queueing-inventory · Queue-dependent replenishment · Markovian arrival process · Phase-type distribution · Matrix-geometric solution

Mathematics Subject Classification 60J28 · 60K25 · 90B05 · 90B22

✉ Serife Ozkar
serife.ozkar@balikesir.edu.tr

Agassi Melikov
amelikov@beu.edu.az

[1] Department of Mathematics, Baku Engineering University, Absheron, Azerbaijan

[2] Department of International Trade and Logistics, Balikesir University, Balikesir, Turkey

1 Introduction

In classical inventory management systems, in-demand items are released directly from the warehouse in accordance with the self-service rule, i.e. in the systems, the time required to serve the customer is zero. So, when the inventory level is zero, customers' arrival causes a result in lost sales or they are served after the replenishment of items. However, a positive time is needed for some necessary procedures to delivery of items in the inventory. The systems in which the service time is positive called queueing-inventory systems (QIS). The first studies in QISs models are Melikov and Molchanov (1992) and Sigman and Simchi-Levi (1992). A detailed survey for the literature of QISs can be examined in Bijvank et al. (2011), Krishnamoorthy et al. (2011, 2021) and Salini et al. (2023).

A distinction should be made between two parts of a traditional QISs: the service facility and the warehouse facility. The service facility contains the server and the buffer for waiting customers, including the service mechanism and sale schemas; the warehouse facility contains the stock for storing inventory items, including the external source(s) and the adopted replenishment policy. According to this framework, there are two types of metrics for assessing the performance of real-world QISs.

Quality of Service (QoS) is a key customer-related factor that directly affects customer satisfaction. QoS metrics (or indicators) are defined using some performance measures, such as loss rate (or loss probability), average queue time (sojourn time), average queue length, server load, etc. Another important issue in QIS is inventory-related factor which is determined by average inventory level, replenishment rate, reorder cycles, average order size, etc. Key issue in the study of QISs models is to balance between customer-related and inventory-related metrics to minimize the system's Expected Total Cost (ETC). In general, this goal can be achieved by using an effective management strategy, typically implemented by (1) using an optimal admission control scheme and (2) implementing an optimal replenishment policy (RP). These approaches are more realistic for achieving the stated goal, since in practice the process of the customer arrival, their service time, warehouse capacity, order fulfillment times, etc. are often not subject to control. In other words, usually only by choosing the either appropriate admission control scheme or replenishment policy it is possible manage ETC and achieve the desired level of performance measures.

Recent papers (Chakravarthy and Melikov 2024; Melikov et al. 2018, 2019; Otten and Daduna 2022; Rasmi et al. 2021; Shajin et al. 2020; Sugapriya et al. 2022, 2023; Varghese and Shajin 2018) have considered QIS models with optimal admission control scheme. In Chakravarthy and Melikov (2024) the reader can find a detailed overview of the works in the first direction. In this study, we focus on the second direction to develop an effective management strategy. There is a significant gap in the study of the concept of optimal RP in QISs. Although previous studies have analyzed some aspects of this problem for single-source QIS. In known studies, the problem of finding the optimal RP is simple reduced to the problem of finding the optimal reorder point when using well-known (s, S)

and (s, Q), $Q = S - s$, policies, where S denotes the maximum size of the system's warehouse and s is reorder point; $s < (S/2)$ for (s, S) policy and $s < S$ for (s, Q) policy.

Note that finding the optimal RP based on both the current queue length and the inventory level has not been well studied, although the relevance of this problem is beyond doubt. We call this type of RP as state-dependent (or queue-dependent) policy. From a scientific point of view, the study of QISs models with state-dependent RPs will enrich the theory of operations research. From a practical point of view, studying the QIS models with state-dependent RPs will allow the system manager to make effective decisions during operational management. Indeed, in case of using state-independent RP, excess stock that have no buyers may arise, which means that the system will waste useless economic costs on storing stock; and vice versa, in this case there may be frequent loss of customers due to lack of stock. In other words, to improve the efficiency of QIS, state-dependent RPs must be applied when ordering inventory.

Let us review the available papers in which single-source QISs models with state-dependent RPs are considered. Melikov and Molchanov (1992) was among the early pioneers to introduce the notion of state-dependent RP within QISs and examined the following model. Threshold s, $0 \le s \le S-1$, where S denotes the maximum capacity of a warehouse, is defined and if the inventory level is greater than s, then no restocking order is sent; otherwise, the queue-dependent randomized replenishment policy is determined as follows: the system orders an inventory of size m, $1 \le m \le S-s$, with probability (w.p.) $\alpha_m(n)$, where n is the number of customers in queue, $0 \le n \le N, N$ denotes the maximum queue capacity. It is assumed that $\sum_{m=1}^{S-s} \alpha_m(n) = 1$ for each n. The authors uses Markov Decision Process (MDP) approach to minimizing the total cost associated with the waiting times and the loss of customers and the holding of stock in the warehouse. The problem is solved by selecting optimal values of the probabilities $\alpha_m(n)$, $1 \le m \le S-s$, $0 \le n \le N$. It is shown that the optimal RP is in the class of nonrandomized (deterministic) policies, i.e. for each n there exist only one $m \in \{1, \ldots, S-s\}$ such that $\alpha_m(n) = 1$, see (Mine and Osaki 1970). In other words, as a result of solving the optimization problem, the optimal size of the order is determined depending on the number of customers in the queue. An approximate method to solving the large scale MDP problem is developed as well. A generalization of this model to the case where the size of items requested by the customers is a random variable was examined in Melikov and Fatalieva (1998). Models of QISs with state-dependent RPs in case *infinity warehouse capacity* (i.e. when $S = \infty$) has been considered in Berman and Kim (1999); Berman and Sapna (2000); He and Jewkes (2000); He et al. (2002a, 2002b) and Kim (2005). In Kim (2005) a QIS model with a finite buffer for waiting of customers is considered. The order size is constant and at each decision point the proposed RP must determine: *Do Not Order* and *Order*. Finding the optimal replenishment policy is a continuous-time MDP, but using the well-known uniformization procedure it is formulated as a discrete-time discounted MDP. The authors show that the optimal RP has a monotone threshold form and develop a procedure that allows finding

the optimal values of the buffer size and order quantity. Similar model with infinite queue was studied in He and Jewkes (2000). Models with zero lead time in the case of a finite queue were studied in Berman and Kim (1999) and Berman and Sapna (2000); similar models in the case of an infinite queue were studied in He et al. (2002a) and He et al. (2002b). These papers also use the MDP approach.

Strategies for timely inventory replenishment and for providing the high levels of QoS often require the use of multiple suppliers (or multi-sources). The focus of this study is on understudied multi-sources QISs *with finite warehouse capacity* (i.e. when $S < \infty$). A literature review of multiple-supplier classical models of inventory management systems (without service stations) and their applications to supply chain management issues is provided by Minner (2003). Based on this review, we conclude that multi-sourcing was prevalent in most business areas during those years and there is every reason to believe that this trend will continue in the future. First of all, having multiple sources allows managers to eliminate the risk of dependence on a single source. Secondly, having multiple suppliers allows managers to minimize the risk of price increases at some sources for various reasons, supply disruptions due to technical disasters, political instability in the region, capacity limitations, fluctuations in lead times, etc.

Unlike classical multi-sources inventory management systems, the models of multi-sources queuing-inventory systems are poorly examined. Recently the models of double-sources QISs are proposed by Melikov et al. (2022a, 2022b) and Melikov et al. (2023). In Melikov et al. (2022a) two separate models of QIS with infinite queues under (s, S) and (s, Q), $Q = S - s$, policies are studied. Inventories can be replenished from two sources with different lead times and inventory delivery costs. If the inventory level drops to the ordering point s, then a regular inventory order is generated to the slow source; if inventory levels fall below a certain threshold value r, where $r < s$, then the system instantly cancels the regular order and an emergency order to the fast source is generated. In Melikov et al. (2022b) a hybrid RP in double-sources QIS is defined as follows: if the inventory level drops to the reorder point s, then a regular order of the fixed volume $Q = S - s$ is generated (i.e. (s, Q)-policy is used) to a slow and cheap source; if the inventory level falls below a certain threshold value r, where $r < s$, then the system instantly cancels the regular order and generates an emergency order to a fast and expensive source where the replenishment quantity should be able to bring the inventory level back to S at the replenishment epoch (i.e. (s, S) policy is used). For all QISs in Melikov et al. (2022a) and Melikov et al. (2022b) the ergodicity conditions are found, their stationary distributions are calculated, and formulas are proposed for finding their performance measures. In addition, the problems of minimizing the total cost of the studied systems are solved by choosing the appropriate values of the ordering point s and threshold value r when using different RPs. In Melikov et al. (2023) similar models with a finite waiting room are investigated.

To our best knowledge, the finding of state-dependent optimal RP in multi-source QIS has been unexplored in the available literature. In this study, we propose a new state-dependent RP in double-source QIS that offering a framework for more efficient inventory management strategies in a real-world situation. The proposed RP is aimed at increasing the replenishment rate through a fast source when the queue

length exceeds a given threshold value. Note that speeding up of inventory items may reduce costs associated with lost sales and customers waiting in queue but it can also increase inventory-holding costs and ordering costs, i.e. inventory management problem becomes more complicating, see (Baek 2024; Jose and Nair 2017) and Rejitha and Jose (2017). This paper explores the cost optimization problem as well.

Our goal is developing the simple, implementable, and yet optimal RP in multi-source QIS. Once the appropriate RP has been developed, the next step is to select the efficient mathematical tool to calculate the performance measures of the QIS under the proposed RP. Here we use the Neuts's matrix-analytic method (*MAM*) (Neuts 1981). The theory and practice of *MAM*, as well as the current state of this theory, can be found in the monographs (Dudin et al. 2020; He 2014; Latouche and Ramaswami 1999) and the two-volume monograph (Chakravarthy 2022a, b).

The essential contributions of our paper are as follows:

- A novel RP in an infinite QIS that based on the switching the order replenishment from a slow rate to a fast rate one depending on the current number in the queue is proposed.
- We consider a model with a Markov arrival process (*MAP*), a phase-type distributed service time and an exponential order replenishment time with different rates for different types of orders.
- Mathematical model of the investigated QIS is formulated as a multi-dimensional Markov chain; a stability condition is obtained and the system's steady-state probabilities and some measures related to system's performance are obtained.
- To be specific, the proposed RP is described based on (s, S) policy while the developed approach can be easily applied for other RPs as well.
- How the performance measures behave as changed the parameters of the system is demonstrated, and the results of the problem of minimizing the total costs are discussed.
- The analysis of the *state-independent* queueing-inventory model is presented as a special case of the proposed model. The differences between the two systems, the state-independent RP and the state-dependent RP, are shown by numerical comparisons in terms of both performance measures and cost optimization.

It is clear that, considering the above listed peculiarities will increase the utility of the model in practice. It is important to note that the approach proposed here can also be applied to finding supply modes in single-source inventory systems with different replenishment capabilities: short lead times but high replenishment costs and long lead times but low replenishment costs. Regular orders can be processed using low-speed transport (e.g. ocean freight), while urgent orders can be processed using high-speed transport (e.g. air freight). In other words, another area of application of the proposed approach is the choice of the supply mode in single-source QISs depending on the current queue length.

The paper is structured as follows. A detailed model description is given in Sect. 2, and model analysis is done in Sect. 3. Various system performance measures

are defined and calculated in Sect. 4. In Sect. 5 the analysis of the state-independent queueing model is presented as a special case of the studied model. Sect. 6 analyses the results of numerical experiments to study behavior of performance measures versus defined thresholds as well as optimization of performance measures. The conclusions are given in the Sect. 7.

2 Model description

We discuss a queueing-inventory system with queue-dependent replenishment policy as demonstrated in Fig. 1.

- In the queueing-inventory system, customers arrive to the system occurs according to a *MAP* with parameters $(\boldsymbol{D}_0, \boldsymbol{D}_1)_{m_1}$, where m_1 is the number of phases, the matrix $\boldsymbol{D}_0$ denotes the transition rates without arrival and the matrix $\boldsymbol{D}_1$ represents the transition rates with arrival. The Markov chain of the *MAP* is managed by the matrix $\boldsymbol{D} = \boldsymbol{D}_0 + \boldsymbol{D}_1$. So, the arrival rate is represented by $\lambda = \boldsymbol{\delta}\boldsymbol{D}_1\boldsymbol{e}$ where $\boldsymbol{\delta}$ is the stationary probability vector of the matrix $\boldsymbol{D}$. The vector satisfies $\boldsymbol{\delta}\boldsymbol{D} = \boldsymbol{0}$ and $\boldsymbol{\delta}\boldsymbol{e} = 1$.
- Once the customers arrive to the system wait in a single line (if the server is busy) and are served in the order of their arrivals. When the inventory level drops to zero, the waiting customers in the queue become impatient. We consider a constant impatient rate. That is, when the inventory level drops to zero, only one customer in the queue becomes impatient independently of the number of customers in the queue and the impatience rate is equal to τ.
- The service times follow a phase-type distribution with parameters $(\boldsymbol{\beta}, \boldsymbol{T})$ of dimension m_2. For later use, the service rate is denoted by $\mu = [\boldsymbol{\beta}(-\boldsymbol{T})^{-1}\boldsymbol{e}]^{-1}$ and $\boldsymbol{T}^0$ is the column vector satisfying $\boldsymbol{T}\boldsymbol{e} + \boldsymbol{T}^0 = \boldsymbol{0}$.
- We assume that an external supplier has different vehicles for the delivery in the system. In other words, orders can be replenished in two ways: the regular (slow) order and the urgent (fast) order. Lead times of the two-type orders differ from each other; i.e., if the regular order is made, then the mean lead time is ν_1^{-1} ; if the urgent order is made, then the mean lead time is ν_2^{-1}. Note that $\nu_2^{-1} < \nu_1^{-1}$. In other words, the urgent order requires a shorter delivery time than the regular

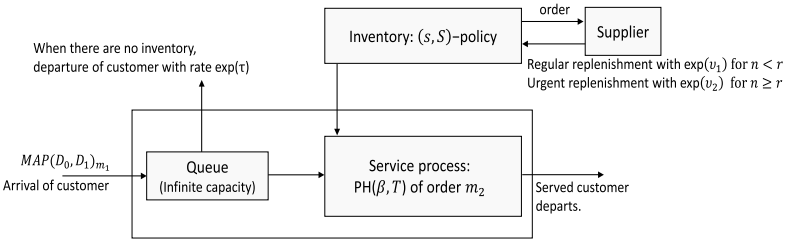

Fig. 1 The queueing-inventory with queue-dependent replenishment policy

order. So, the supplies via an urgent order are more expensive, expectedly. The lead times follow an exponential distribution.

- The queue-dependent replenishment policy with different lead times based on "Up to S" scheme (or (s, S)-policy) is studied. The ordering point s, $0 < s < S$, is defined as follows. It means that if the inventory level is greater than s, then no restocking order is sent; otherwise, the replenishment will bring the inventory to its fullest, i.e., when the replenishment occurs the inventory level will increase to S no matter what the inventory level at the moment.
- We define the threshold value $r, r > 0$, for the number of the customers in the system. So, the queue-dependent replenishment policy is defined as follows: if the inventory level drops to s and the number of the customers in the system is less than r, then the regular order is made; when the inventory level drops to s and the number of the customers in the system is more or equal r, then the regular order is instantly canceled and the urgent order is sent.

3 Steady-state analysis

We investigate the steady-state analysis of the queueing-inventory system. The process $\{(N(t), I(t), J(t), K(t)), t \geq 0\}$ is a continuous-time Markov chain ($CTMC$). That is, at time t, $N(t)$ is the number of customers in the system; $I(t)$ is the inventory level in the system; $J(t)$ is the phase of the service process; and $K(t)$ is the phase of the arrival process. The state space of the system is given by

$$\begin{aligned}\Omega = &\{(0, i, k) : 0 \leq i \leq S,\ k = 1, \ldots, m_1\} \bigcup \\ &\{(n, i, j, k) : n > 0,\ 0 \leq i \leq S,\ j = 1, \ldots, m_2,\ k = 1, \ldots, m_1\}.\end{aligned}$$

The level $\{(0, i, k) : 0 \leq i \leq S,\ k = 1,..., m_1\}$ with dimension $m_1(S + 1)$ conforms to the case when there are no customers in the system, the inventory level is i, and the arrival process is in one of m_1 phases. The level $\{(n, i, j, k) : n > 0,\ 0 \leq i \leq S,\ j = 1, ..., m_2, k = 1,..., m_1\}$ with dimension $m_1m_2(S + 1)$ conforms to the case when there are n customers in the system, the inventory level is i, the service process and the arrival process are in one of m_2 phases and in one of m_1 phases, respectively.

The generator matrix of the system in QBD structure is given by

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{B}_0 & \boldsymbol{A}_0 & & & & & \\ \boldsymbol{C}_0 & \boldsymbol{B}_1 & \boldsymbol{A} & & & & \\ & \boldsymbol{C} & \boldsymbol{B}_1 & \boldsymbol{A} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \boldsymbol{C} & \boldsymbol{B}_1 & \boldsymbol{A} & \\ & & & & \boldsymbol{C} & \boldsymbol{B}_2 & \boldsymbol{A} \\ & & & & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{1}$$

The matrices $\boldsymbol{A}_0$ and $\boldsymbol{A}$ have dimensions $m_1(S + 1) \times m_1m_2(S + 1)$ and $m_1m_2(S + 1) \times m_1m_2(S + 1)$, respectively.

$$\boldsymbol{A}_0 = \begin{pmatrix} \boldsymbol{\beta} \otimes \boldsymbol{D}_1 & & & \\ & \boldsymbol{\beta} \otimes \boldsymbol{D}_1 & & \\ & & \ddots & \\ & & & \boldsymbol{\beta} \otimes \boldsymbol{D}_1 \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{I} \otimes \boldsymbol{D}_1 & & & \\ & \boldsymbol{I} \otimes \boldsymbol{D}_1 & & \\ & & \ddots & \\ & & & \boldsymbol{I} \otimes \boldsymbol{D}_1 \end{pmatrix}.$$

The matrices $\boldsymbol{C}_0$ and $\boldsymbol{C}$ have dimensions $m_1 m_2(S+1) \times m_1(S+1)$ and $m_1 m_2(S+1) \times m_1 m_2(S+1)$, respectively.

$$\boldsymbol{C}_0 = \begin{pmatrix} (\boldsymbol{e} \otimes \tau \boldsymbol{I}) & & & \\ (\boldsymbol{T}^0 \otimes \boldsymbol{I}) & 0 & & \\ & \ddots & \ddots & \\ & & (\boldsymbol{T}^0 \otimes \boldsymbol{I}) & 0 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} \tau \boldsymbol{I} & & & & \\ (\boldsymbol{T}^0 \boldsymbol{\beta} \otimes \boldsymbol{I}) & 0 & & & \\ & (\boldsymbol{T}^0 \boldsymbol{\beta} \otimes \boldsymbol{I}) & 0 & & \\ & & \ddots & \ddots & \\ & & & (\boldsymbol{T}^0 \boldsymbol{\beta} \otimes \boldsymbol{I}) & 0 \end{pmatrix}.$$

The matrices $\boldsymbol{B}_0$ and $\boldsymbol{B}_l$, $l = 1, 2$, have dimensions $m_1(S+1) \times m_1(S+1)$ and $m_1 m_2(S+1) \times m_1 m_2(S+1)$, respectively.

$$\boldsymbol{B}_0 = \begin{pmatrix} \boldsymbol{D}_0 - \nu_1 \boldsymbol{I} & & & & & \nu_1 \boldsymbol{I} \\ & \boldsymbol{D}_0 - \nu_1 \boldsymbol{I} & & & & \nu_1 \boldsymbol{I} \\ & & \ddots & & & \vdots \\ & & & \boldsymbol{D}_0 - \nu_1 \boldsymbol{I} & & \nu_1 \boldsymbol{I} \\ & & & & \boldsymbol{D}_0 & \\ & & & & & \ddots \\ & & & & & \boldsymbol{D}_0 \end{pmatrix},$$

$$\boldsymbol{B}_l = \begin{pmatrix} (\boldsymbol{I} \otimes \boldsymbol{D}_0) - (\tau + \nu_l)\boldsymbol{I} & & & & & & \nu_l \boldsymbol{I} \\ & (\boldsymbol{T} \oplus \boldsymbol{D}_0) - \nu_l \boldsymbol{I} & & & & & \nu_l \boldsymbol{I} \\ & & \ddots & & & & \vdots \\ & & & (\boldsymbol{T} \oplus \boldsymbol{D}_0) - \nu_l \boldsymbol{I} & & & \nu_l \boldsymbol{I} \\ & & & & (\boldsymbol{T} \oplus \boldsymbol{D}_0) & & \\ & & & & & \ddots & \\ & & & & & & (\boldsymbol{T} \oplus \boldsymbol{D}_0) \end{pmatrix}$$

3.1 Stability condition

We give the invariant vector of the generator matrix $\boldsymbol{F} = \boldsymbol{A} + \boldsymbol{B}_2 + \boldsymbol{C}$ by the vector $\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_s, \boldsymbol{\pi}_{s+1}, \cdots, \boldsymbol{\pi}_S)$. Each vector part $\boldsymbol{\pi}_i$ of dimension mn corresponds that the inventory level is i, the service process is in one of the n phases, and the arrival process is in one of the m phases. So, the vector $\boldsymbol{\pi}$ satisfies

$$\boldsymbol{\pi F} = \boldsymbol{0} \text{ and } \boldsymbol{\pi e} = 1. \tag{2}$$

Theorem 1 *The queueing-inventory system under study is stable if and only if the following condition is satisfied*

$$\lambda < \mu(1 - \boldsymbol{\pi}_0 \boldsymbol{e}) + \tau \boldsymbol{\pi}_0 \boldsymbol{e}. \tag{3}$$

Proof The system under study is a *QBD* process. Therefore, it is stable *if and only if* $\boldsymbol{\pi A e} < \boldsymbol{\pi C e}$ (See in Neuts 1981). That is,

$$\sum_{i=0}^{S} \boldsymbol{\pi}_i(\boldsymbol{I}_{m_2} \otimes \boldsymbol{D}_1)\boldsymbol{e} \;<\; \tau \boldsymbol{\pi}_0 \boldsymbol{e} + \sum_{i=1}^{S} \boldsymbol{\pi}_i(\boldsymbol{T}^0\boldsymbol{\beta} \otimes \boldsymbol{I}_{m_1})\boldsymbol{e}. \tag{4}$$

Firstly, we rewrite the steady-state equations in (2) as follows.

$$\begin{aligned}
\boldsymbol{\pi}_0[(\boldsymbol{I} \otimes \boldsymbol{D}_1) + (\boldsymbol{I} \otimes \boldsymbol{D}_0) - \nu_2 \boldsymbol{I}] + \boldsymbol{\pi}_1(\boldsymbol{T}^0\boldsymbol{\beta} \otimes \boldsymbol{I}) &= \boldsymbol{0}, \\
\boldsymbol{\pi}_i[(\boldsymbol{I} \otimes \boldsymbol{D}_1) + (\boldsymbol{T} \oplus \boldsymbol{D}_0) - \nu_2 \boldsymbol{I}] + \boldsymbol{\pi}_{i+1}(\boldsymbol{T}^0\boldsymbol{\beta} \otimes \boldsymbol{I}) &= \boldsymbol{0}, \; 1 \le i \le s, \\
\boldsymbol{\pi}_i[(\boldsymbol{I} \otimes \boldsymbol{D}_1) + (\boldsymbol{T} \oplus \boldsymbol{D}_0)] + \boldsymbol{\pi}_{i+1}(\boldsymbol{T}^0\boldsymbol{\beta} \otimes \boldsymbol{I}) &= \boldsymbol{0}, \; s+1 \le i \le S-1, \\
[\boldsymbol{\pi}_0 + \cdots + \boldsymbol{\pi}_s]\nu_2 \boldsymbol{I} + \boldsymbol{\pi}_S[(\boldsymbol{I} \otimes \boldsymbol{D}_1) + (\boldsymbol{T} \oplus \boldsymbol{D}_0)] &= \boldsymbol{0},
\end{aligned} \tag{5}$$

with the normalizing condition

$$\sum_{i=0}^{S} \boldsymbol{\pi}_i \boldsymbol{e} = 1. \tag{6}$$

By adding the equations in (5), the following equation is achieved

$$\boldsymbol{\pi}_0(\boldsymbol{I}_{m_2} \otimes \boldsymbol{D}) + \sum_{i=1}^{S} \boldsymbol{\pi}_i\left[(\boldsymbol{T} + \boldsymbol{T}^0\boldsymbol{\beta}) \oplus \boldsymbol{D}\right] = \boldsymbol{0}. \tag{7}$$

Post-multiplying the equation (7) by $(\boldsymbol{e}_{m_2} \otimes \boldsymbol{I}_{m_1})$ and using the rate $\lambda = \boldsymbol{\delta D}_1 \boldsymbol{e}$, and then by the condition in (6), we get the left side of the inequality in (4) as follows.

$$\sum_{i=0}^{S} \boldsymbol{\pi}_i\left(\boldsymbol{e}_{m_2} \otimes \boldsymbol{D_1 e}_{m_1}\right) = \lambda \sum_{i=0}^{S} \boldsymbol{\pi}_i \boldsymbol{e} = \lambda.$$

Post-multiplying the equation (7) by $(\boldsymbol{I}_{m_2} \otimes \boldsymbol{e}_{m_1})$ and using the rate $\mu = 1/[\boldsymbol{\beta}(-\boldsymbol{T})^{-1}\boldsymbol{e}]$ and the condition in (6), we get

$$\sum_{i=1}^{S} \boldsymbol{\pi}_i\left(\boldsymbol{T}^0\boldsymbol{\beta} \otimes \boldsymbol{I}_{m_1}\right)\boldsymbol{e} = \mu \sum_{i=1}^{S} \boldsymbol{\pi}_i \boldsymbol{e} = \mu(1 - \boldsymbol{\pi}_0 \boldsymbol{e}).$$

This result gives the right-side of the inequality in (4) as $\mu(1 - \boldsymbol{\pi}_0\boldsymbol{e}) + \tau\boldsymbol{\pi}_0\boldsymbol{e}$. In this way, the proof for Theorem 1 is completed. □

Note 1. The established stability condition in (3) has a probabilistic meaning. The condition denotes that the arrival rate of customers to the system must be less than the total rate of impatient rate (leaving the system because of no inventory) and the service rate of customers.

3.2 The steady-state probability vector

We give the steady-state probability vector for the generator matrix $\boldsymbol{Q}$ in (1) by $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_r, \boldsymbol{x}_{r+1}, \cdots)$. So, the vector $\boldsymbol{x}$ satisfies the following equations.

$$\boldsymbol{x}\,\boldsymbol{Q} = \boldsymbol{0} \text{ and } \boldsymbol{x}\,\boldsymbol{e} = 1. \tag{8}$$

The vector $\boldsymbol{x}_0$ with dimension $m_1(S+1)$ is partitioned into the vectors as $\boldsymbol{x}_0 = [\boldsymbol{x}_0(0), \boldsymbol{x}_0(1), \cdots, \boldsymbol{x}_0(S)]$. In here, the dimension of each vector is m_1. The vector $\boldsymbol{x}_0(i)$ denotes the probability that there are no customers in the system, the inventory level is i, $0 \le i \le S$, and the arrival process is in one of the m_1 phases.

$m_1 m_2(S+1)$ dimensional vector $\boldsymbol{x}_n$, $n \ge 1$, is also partitioned into the vectors as $\boldsymbol{x}_n = [\boldsymbol{x}_n(0), \boldsymbol{x}_n(1), \cdots, \boldsymbol{x}_n(S)]$. The dimension of each vector is $m_1 m_2$. The vector $\boldsymbol{x}_n(i)$ represents the probability that there are n customers in the system, the inventory level is i, $0 \le i \le S$, the service process is in one of the m_2 phases and the arrival process is in one of the m_1 phases.

To obtain the steady-state probability vector $\boldsymbol{x}$, we use the following Theorem 2 as a solution to the equations in (8). The theorem is a direct consequence of Neuts' result on *QBD* processes (see, Neuts 1981).

Theorem 2 *Under the stability condition in Theorem 1 the steady-state probability vector* $\boldsymbol{x}$ *is achieved by solving the following system of linear equations*

$$\boldsymbol{x}_0\boldsymbol{B}_0 + \boldsymbol{x}_1\boldsymbol{C}_0 = \boldsymbol{0}, \tag{9}$$

$$\boldsymbol{x}_0\boldsymbol{A}_0 + \boldsymbol{x}_1\boldsymbol{B}_1 + \boldsymbol{x}_2\boldsymbol{C} = \boldsymbol{0}, \tag{10}$$

$$\boldsymbol{x}_{n-1}\boldsymbol{A} + \boldsymbol{x}_n\boldsymbol{B}_1 + \boldsymbol{x}_{n+1}\boldsymbol{C} = \boldsymbol{0}, \quad 2 \le n \le r-2 \tag{11}$$

$$\boldsymbol{x}_{r-2}\boldsymbol{A} + \boldsymbol{x}_{r-1}[\boldsymbol{B}_1 + \boldsymbol{R}\boldsymbol{C}] = \boldsymbol{0}, \tag{12}$$

$$\boldsymbol{x}_0\boldsymbol{e} + \sum_{n=1}^{r-2} \boldsymbol{x}_n\boldsymbol{e} + \boldsymbol{x}_{r-1}(\boldsymbol{I} - \boldsymbol{R})^{-1}\boldsymbol{e} = 1. \tag{13}$$

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

$$\boldsymbol{R}^2\boldsymbol{C} + \boldsymbol{R}\boldsymbol{B}_2 + \boldsymbol{A} = \boldsymbol{0}. \tag{14}$$

The QBD structure of the generator matrix in (1) *yields a modified matrix-geometric solution. So, the non-boundary states* ($n \ge r$) *are given by*

$$\boldsymbol{x}_{n+r} = \boldsymbol{x}_{r-1}\boldsymbol{R}^{n+1}, \quad n \ge 0. \tag{15}$$

4 Performance measures

Performance measures related to the queueing-inventory system with queue-dependent replenishment policy are listed.

1. **Measures related to customer**
   The average number of customers in the system:

$$L_{av} = \sum_{n=1}^{\infty} n\, \boldsymbol{x}_n \boldsymbol{e} = \sum_{n=1}^{r-1} n\, \boldsymbol{x}_n \boldsymbol{e} + \boldsymbol{x}_{r-1}\left[r\boldsymbol{R}(\boldsymbol{I}-\boldsymbol{R})^{-1} + \boldsymbol{R}^2(\boldsymbol{I}-\boldsymbol{R})^{-2}\right]\boldsymbol{e}. \tag{16}$$

   The probability of customers leaving the system when there is no inventory:

$$P_{lost} = \sum_{n=1}^{\infty} \boldsymbol{x}_n(0)\boldsymbol{e}_{m_1 m_2}. \tag{17}$$

2. **Measures related to inventory**
   The average number of items in the inventory

$$I_{av} = \sum_{i=1}^{S} i\, \boldsymbol{x}_0(i)\boldsymbol{e}_{m_1} + \sum_{n=1}^{\infty}\sum_{i=1}^{S} i\, \boldsymbol{x}_n(i)\boldsymbol{e}_{m_1 m_2}. \tag{18}$$

   The average volume of deliveries via the regular order

$$V_r = \sum_{i=S-s}^{S} i\, \boldsymbol{x}_0(S-i)\boldsymbol{e}_{m_1} + \sum_{n=1}^{r-1}\sum_{i=S-s}^{S} i\, \boldsymbol{x}_n(S-i)\boldsymbol{e}_{m_1 m_2}. \tag{19}$$

   The average volume of deliveries via the urgent order

$$V_u = \sum_{n=r}^{\infty}\sum_{i=S-s}^{S} i\, \boldsymbol{x}_n(S-i)\boldsymbol{e}_{m_1 m_2}. \tag{20}$$

3. **Measures related to replenishment**
   The average intensity of regular orders

$$RR_r = \sum_{k=1}^{r-1} \boldsymbol{x}_n(s+1)(\boldsymbol{T}^0 \otimes \boldsymbol{I}_{m_1})\boldsymbol{e}_{m_1}. \tag{21}$$

   The average intensity of urgent orders

$$RR_u = \sum_{k=r}^{\infty} \boldsymbol{x}_n(s+1)(\boldsymbol{T}^0 \otimes \boldsymbol{I}_{m_1})\boldsymbol{e}_{m_1}. \tag{22}$$

5 State-independent case of the studied queueing-inventory model

In this section, the queue-independent replenishment policy is considered for the studied queueing-inventory model. So, we ignore the threshold value r defined for the number of the customers in the system. Regardless of the number of the customers in the system, an order is placed when the inventory number drops to s. The lead times follow an exponential distribution with parameter ν. In other words, the replenishment rates of the regular orders and the urgent orders, previously considered as ν_1 and ν_2, respectively, are assumed to be equal in this model ($\nu = \nu_1 = \nu_2$). The urgent order situation, which is given depending on the threshold value r, is ignored. Except for the replenishment policy, all assumptions of the model defined in Sect. 2 are valid for the state-independent queueing-inventory model studied here.

Due to the replenishment rate, only the matrices in the main diagonal of the generator matrix $\boldsymbol{Q}$ in (1) change while the other matrices remain the same. The new generating matrix is given follows.

$$\boldsymbol{Q}_1 = \begin{pmatrix} \boldsymbol{K}_0 & \boldsymbol{A}_0 & & & \\ \boldsymbol{C}_0 & \boldsymbol{K} & \boldsymbol{A} & & \\ & \boldsymbol{C} & \boldsymbol{K} & \boldsymbol{A} & \\ & & \ddots & \ddots & \ddots \end{pmatrix}, \tag{23}$$

with

$$\boldsymbol{K}_0 = \begin{pmatrix} \boldsymbol{D}_0 - \nu\boldsymbol{I} & & & & & \nu\boldsymbol{I} \\ & \boldsymbol{D}_0 - \nu\boldsymbol{I} & & & & \nu\boldsymbol{I} \\ & & \ddots & & & \vdots \\ & & & \boldsymbol{D}_0 - \nu\boldsymbol{I} & & \nu\boldsymbol{I} \\ & & & & \boldsymbol{D}_0 & \\ & & & & & \ddots \\ & & & & & \boldsymbol{D}_0 \end{pmatrix},$$

$$\boldsymbol{K} = \begin{pmatrix} (\boldsymbol{I} \otimes \boldsymbol{D}_0) - (\tau + \nu)\boldsymbol{I} & & & & & & \nu\boldsymbol{I} \\ & (\boldsymbol{T} \oplus \boldsymbol{D}_0) - \nu\boldsymbol{I} & & & & & \nu\boldsymbol{I} \\ & & \ddots & & & & \vdots \\ & & & (\boldsymbol{T} \oplus \boldsymbol{D}_0) - \nu\boldsymbol{I} & & & \nu\boldsymbol{I} \\ & & & & (\boldsymbol{T} \oplus \boldsymbol{D}_0) & & \\ & & & & & \ddots & \\ & & & & & & (\boldsymbol{T} \oplus \boldsymbol{D}_0) \end{pmatrix}.$$

We give the invariant vector of the matrix $\boldsymbol{F}_1 = \boldsymbol{A} + \boldsymbol{K} + \boldsymbol{C}$ by the vector $\hat{\boldsymbol{\pi}} = (\hat{\boldsymbol{\pi}}_0, \hat{\boldsymbol{\pi}}_1, \cdots, \hat{\boldsymbol{\pi}}_s, \hat{\boldsymbol{\pi}}_{s+1}, \cdots, \hat{\boldsymbol{\pi}}_S)$. The vector $\hat{\boldsymbol{\pi}}$ satisfies

$$\hat{\boldsymbol{\pi}}\boldsymbol{F}_1 = \boldsymbol{0} \text{ and } \hat{\boldsymbol{\pi}}\boldsymbol{e} = 1. \tag{24}$$

The stability condition of the state-independent queueing-inventory model is given in the following Theorem 3. We note that the stability condition in (25) is similar to one in Theorem 1 (consider the vector $\hat{\boldsymbol{\pi}}_0$ instead of $\boldsymbol{\pi}_0$).

Theorem 3 *The queueing-inventory system with state independent is stable if and only if the following condition is satisfied*

$$\lambda < \mu(1 - \hat{\boldsymbol{\pi}}_0\boldsymbol{e}) + \tau\hat{\boldsymbol{\pi}}_0\boldsymbol{e}. \tag{25}$$

Proof The state-independent queueing-inventory system in this section is also a *QBD* process. So, the system is stable *if and only if* $\hat{\boldsymbol{\pi}}\boldsymbol{A}\boldsymbol{e} < \hat{\boldsymbol{\pi}}\boldsymbol{C}\boldsymbol{e}$ (See in Neuts 1981). That is,

$$\sum_{i=0}^{S} \hat{\boldsymbol{\pi}}_i(\boldsymbol{I}_{m_2} \otimes \boldsymbol{D}_1)\boldsymbol{e} \;<\; \tau\hat{\boldsymbol{\pi}}_0\boldsymbol{e} + \sum_{i=1}^{S} \hat{\boldsymbol{\pi}}_i(\boldsymbol{T}^0\boldsymbol{\beta} \otimes \boldsymbol{I}_{m_1})\boldsymbol{e}. \tag{26}$$

Firstly, we rewrite the steady-state equations in (24) as follows.

$$\begin{aligned}
\hat{\boldsymbol{\pi}}_0[(\boldsymbol{I} \otimes \boldsymbol{D}_1) + (\boldsymbol{I} \otimes \boldsymbol{D}_0) - \nu\boldsymbol{I}] + \hat{\boldsymbol{\pi}}_1(\boldsymbol{T}^0\boldsymbol{\beta} \otimes \boldsymbol{I}) &= \boldsymbol{0},\\
\hat{\boldsymbol{\pi}}_i[(\boldsymbol{I} \otimes \boldsymbol{D}_1) + (\boldsymbol{T} \oplus \boldsymbol{D}_0) - \nu\boldsymbol{I}] + \hat{\boldsymbol{\pi}}_{i+1}(\boldsymbol{T}^0\boldsymbol{\beta} \otimes \boldsymbol{I}) &= \boldsymbol{0},\ 1 \le i \le s,\\
\hat{\boldsymbol{\pi}}_i[(\boldsymbol{I} \otimes \boldsymbol{D}_1) + (\boldsymbol{T} \oplus \boldsymbol{D}_0)] + \hat{\boldsymbol{\pi}}_{i+1}(\boldsymbol{T}^0\boldsymbol{\beta} \otimes \boldsymbol{I}) &= \boldsymbol{0},\ s+1 \le i \le S-1,\\
[\hat{\boldsymbol{\pi}}_0 + \cdots + \hat{\boldsymbol{\pi}}_s]\nu\boldsymbol{I} + \hat{\boldsymbol{\pi}}_S[(\boldsymbol{I} \otimes \boldsymbol{D}_1) + (\boldsymbol{T} \oplus \boldsymbol{D}_0)] &= \boldsymbol{0},\\
\text{with the normalizing condition } \sum_{i=0}^{S} \hat{\boldsymbol{\pi}}_i\boldsymbol{e} &= 1.
\end{aligned} \tag{27}$$

By adding the equations in (27), we obtain the following equation.

$$\hat{\boldsymbol{\pi}}_0(\boldsymbol{I}_{m_2} \otimes \boldsymbol{D}) + \sum_{i=1}^{S} \hat{\boldsymbol{\pi}}_i\big[(\boldsymbol{T} + \boldsymbol{T}^0\boldsymbol{\beta}) \oplus \boldsymbol{D}\big] = \boldsymbol{0}. \tag{28}$$

The equation in (28) is multiplied by $(\boldsymbol{e}_{m_2} \otimes \boldsymbol{I}_{m_1})$ and then we get the left side of the inequality in (26) by using the arrival rate $\lambda = \boldsymbol{\delta}\boldsymbol{D}_1\boldsymbol{e}$ and the normalizing condition in (27) as follows.

$$\sum_{i=0}^{S} \hat{\boldsymbol{\pi}}_i\big(\boldsymbol{e}_{m_2} \otimes \boldsymbol{D}_1\boldsymbol{e}_{m_1}\big) = \lambda \sum_{i=0}^{S} \hat{\boldsymbol{\pi}}_i\boldsymbol{e} = \lambda.$$

After the equation in (28) is multiplied by $(\boldsymbol{I}_{m_2} \otimes \boldsymbol{e}_{m_1})$, we obtain the following result by using the service rate $\mu = 1/[\boldsymbol{\beta}(-\boldsymbol{T})^{-1}\boldsymbol{e}]$ and the normalizing condition in (27).

$$\sum_{i=1}^{S} \hat{\boldsymbol{\pi}}_i\big(\boldsymbol{T}^0\boldsymbol{\beta} \otimes \boldsymbol{I}_{m_1}\big)\boldsymbol{e} = \mu \sum_{i=1}^{S} \hat{\boldsymbol{\pi}}_i\boldsymbol{e} = \mu(1 - \hat{\boldsymbol{\pi}}_0\boldsymbol{e}).$$

The result gives the right-side of the inequality in (26) as $\mu(1 - \boldsymbol{\pi}_0\boldsymbol{e}) + \tau\boldsymbol{\pi}_0\boldsymbol{e}$. So, the proof for Theorem 3 is completed. □

We give the steady-state probability vector for the generator matrix in (23) by $\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, \cdots)$. So, the vector $\boldsymbol{y}$ satisfies the following equations.

$$\boldsymbol{y}\,\boldsymbol{Q}_1 = \boldsymbol{0} \;\text{ and }\; \boldsymbol{y}\,\boldsymbol{e} = 1. \tag{29}$$

The following Theorem 4 is used to obtain the steady-state probability vector $\boldsymbol{y}$. The theorem is a direct consequence of Neuts' result on *QBD* processes (see, Neuts 1981).

Theorem 4 *Under the stability condition in Theorem 3 the steady-state probability vector* $\boldsymbol{y}$ *is achieved by solving the following system of linear equations*

$$\boldsymbol{y}_0\boldsymbol{K}_0 + \boldsymbol{y}_1\boldsymbol{C}_0 = \boldsymbol{0}, \tag{30}$$

$$\boldsymbol{y}_0\boldsymbol{A}_0 + \boldsymbol{y}_1[\boldsymbol{K} + \boldsymbol{R}_1\boldsymbol{C}] = \boldsymbol{0}, \tag{31}$$

$$\boldsymbol{y}_0\boldsymbol{e} + \boldsymbol{y}_1(\boldsymbol{I} - \boldsymbol{R}_1)^{-1}\boldsymbol{e} = 1. \tag{32}$$

where the matrix $\boldsymbol{R}_1$ *is the minimal nonnegative solution to the following matrix quadratic equation*

$$\boldsymbol{R}_1^2\boldsymbol{C} + \boldsymbol{R}_1\boldsymbol{K} + \boldsymbol{A} = \boldsymbol{0}. \tag{33}$$

The states ($n \geq 2$) *are given by*

$$\boldsymbol{y}_n = \boldsymbol{y}_1\boldsymbol{R}_1^{n-1}, \quad n \geq 2. \tag{34}$$

Note 2. Due to the structures of the generator matrices, the geometric structure going to infinity (*the solution for the non-boundary states*) is captured in the state-dependent model at the case $n \geq r$ in (15) while it is captured in the state-independent model at the case $n \geq 2$ in (34). So, at state-dependent model in Sect. 3, the structure of the generator matrix $\boldsymbol{Q}$ in (1) yields modified matrix-geometric solution as given Theorem 2. On the other hand, at state-independent model in this section, the structure of the generator matrix $\boldsymbol{Q}_1$ in (23) yields original matrix-geometric solution as given Theorem 4.

Finally, performance measures related to the queueing-inventory system with queue-independent replenishment policy are listed.

1. **Measures related to customer**

The average number of customers in the system:

$$\hat{L}_{av} = \sum_{n=1}^{\infty} n\,\boldsymbol{y}_n\boldsymbol{e} = \boldsymbol{y}_1(\boldsymbol{I} - \boldsymbol{R}_1)^{-2}\boldsymbol{e}. \tag{35}$$

The probability of customers leaving the system when there is no inventory:

$$\hat{P}_{lost} = \sum_{n=1}^{\infty} \boldsymbol{y}_n(0)\boldsymbol{e}_{m_1m_2}. \tag{36}$$

2. **Measures related to inventory**
The average number of items in the inventory

$$\hat{I}_{av} = \sum_{i=1}^{S} i\, \boldsymbol{y}_0(i)\boldsymbol{e}_{m_1} + \sum_{n=1}^{\infty}\sum_{i=1}^{S} i\, \boldsymbol{y}_n(i)\boldsymbol{e}_{m_1 m_2}. \tag{37}$$

The average volume of deliveries

$$\hat{V} = \sum_{i=S-s}^{S} i\, \boldsymbol{y}_0(S-i)\boldsymbol{e}_{m_1} + \sum_{n=1}^{\infty}\sum_{i=S-s}^{S} i\, \boldsymbol{y}_n(S-i)\boldsymbol{e}_{m_1 m_2}. \tag{38}$$

3. **Measures related to replenishment**
The average intensity of orders

$$\hat{R}_R = \sum_{k=1}^{\infty} \boldsymbol{y}_n(s+1)(\boldsymbol{T}^0 \otimes \boldsymbol{I}_{m_1})\boldsymbol{e}_{m_1}. \tag{39}$$

6 Numerical illustration

The behavior of the some performance measures and optimum inventory policy under various arrival processes and service time distributions are discussed.

For the arrival process of the system, the following values of the matrices $\boldsymbol{D}_0$ and $\boldsymbol{D}_1$ are considered. All processes has the same mean of 1. However, each of them is qualitatively different. Namely, the values of the standard deviation for the inter-arrival times of ER-A, EX-A, HE-A NC-A and PC-A are, respectively, 1, 1.41421, 3.17451, 1.99336, and 1.99336. We note that the values of the standard deviation are given according to ER-A. The *MAP* processes are normalized to have a specific arrival rate λ. For two sequential inter-arrival times, the process NC-A has a negative correlation with a value of -0.4889 and the process PC-A has a positive correlation with a value of 0.4889. There are no correlation in the other arrival processes.

Table 1 The values of the system parameters

As It Is Varied	It Is Fixed
The arrival rate: λ	$r = 4, \mu = 4, \nu_1 = 1, \nu_2 = 2.5, \tau = 5$
The service rate: μ	$r = 4, \lambda = 1.6, \nu_1 = 1, \nu_2 = 2.5, \tau = 5$
The impatient rate: τ	$r = 4, \lambda = 3.8, \mu = 4, \nu_1 = 1, \nu_2 = 2.5$
The threshold point to switch order mode: r	$\lambda = 3.8, \mu = 4, \nu_1 = 1, \nu_2 = 2.5, \tau = 5$
The replenishment rate for regular order: ν_1	$r = 4, \lambda = 3.8, \mu = 4, \nu_2 = 3, \tau = 5$
The replenishment rate for urgent order: ν_2	$r = 4, \lambda = 3.8, \mu = 4, \nu_1 = 1, \tau = 5$

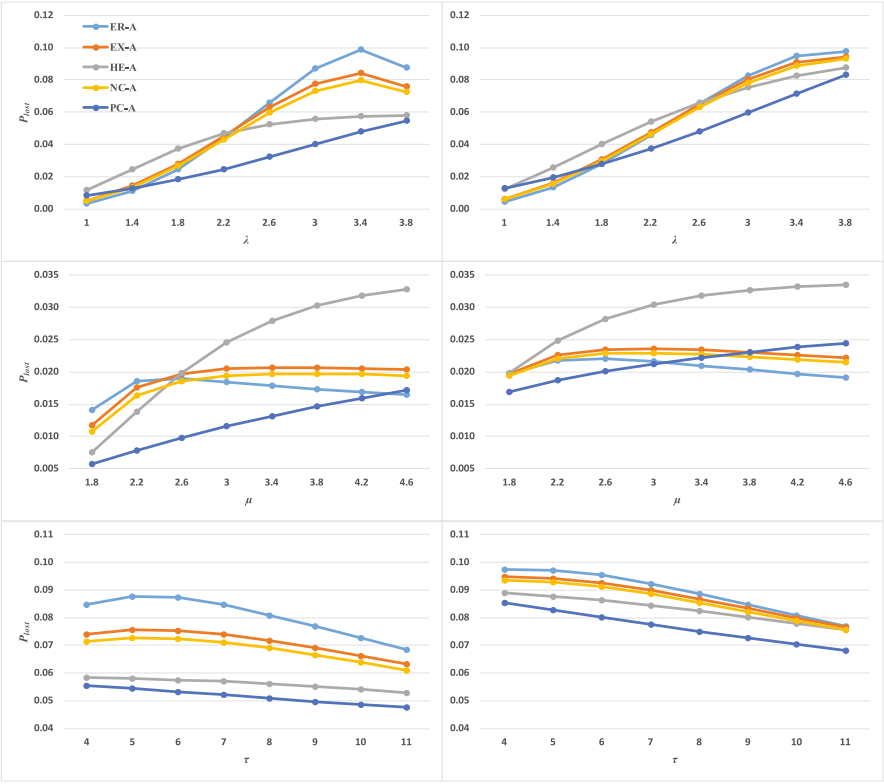

Fig. 2 P_{lost} vs λ, μ, τ for ER-S (left side) and HE-S (right side)

Erlang distribution (ER-A): $\boldsymbol{D}_0 = \begin{pmatrix} -2 & 2 \\ 0 & -2 \end{pmatrix}$, $\boldsymbol{D}_1 = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}$.

Exponential distribution (EX-A): $\boldsymbol{D}_0 = \begin{pmatrix} -1 \end{pmatrix}$, $\boldsymbol{D}_1 = \begin{pmatrix} 1 \end{pmatrix}$.

Hyperexponential distribution (HE-A):

$$\boldsymbol{D}_0 = \begin{pmatrix} -1.9 & 0 \\ 0 & -0.19 \end{pmatrix}, \quad \boldsymbol{D}_1 = \begin{pmatrix} 1.71 & 0.19 \\ 0.171 & 0.019 \end{pmatrix}.$$

MAP with negative correlation (NC-A):

$$\boldsymbol{D}_0 = \begin{pmatrix} -1.00222 & 1.00222 & 0 \\ 0 & -1.00222 & 0 \\ 0 & 0 & -225.75 \end{pmatrix}, \quad \boldsymbol{D}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0.01002 & 0 & 0.9922 \\ 223.4925 & 0 & 2.2575 \end{pmatrix}.$$

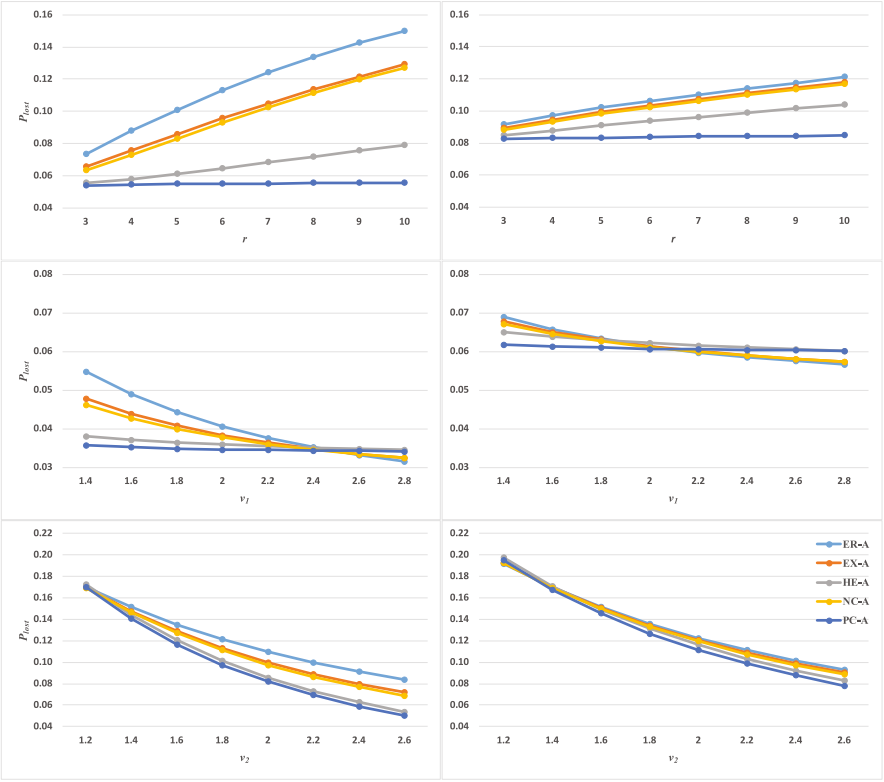

Fig. 3 P_{lost} vs r, ν_1, ν_2 for ER-S (left side) and HE-S (right side)

MAP with positive correlation (PC-A):

$$\boldsymbol{D}_0 = \begin{pmatrix} -1.00222 & 1.00222 & 0 \\ 0 & -1.00222 & 0 \\ 0 & 0 & -225.75 \end{pmatrix}, \quad \boldsymbol{D}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0.9922 & 0 & 0.01002 \\ 2.2575 & 0 & 223.4925 \end{pmatrix}.$$

For the service times, the phase-type distributions with parameter $(\boldsymbol{\beta}, \boldsymbol{T})$ are considered. Each of the distributions has the same mean of 1, but qualitatively different. The values of the standard deviation of ER-S, EX-S and HE-S are, respectively, 0.70711, 1 and 2.24472. The distributions are normalized at a specific value for the service rate μ.

Erlang distribution (ER-S): $\boldsymbol{\beta} = \begin{pmatrix} 1, & 0 \end{pmatrix}, \quad \boldsymbol{T} = \begin{pmatrix} -2 & 2 \\ 0 & -2 \end{pmatrix}.$

Exponential distribution (EX-S): $\boldsymbol{\beta} = \begin{pmatrix} 1 \end{pmatrix}, \quad \boldsymbol{T} = \begin{pmatrix} -1 \end{pmatrix}.$

Hyperexponential distribution (HE-S): $\boldsymbol{\beta} = \begin{pmatrix} 0.9, & 0.1 \end{pmatrix}, \quad \boldsymbol{T} = \begin{pmatrix} -1.9 & 0 \\ 0 & -0.19 \end{pmatrix}.$

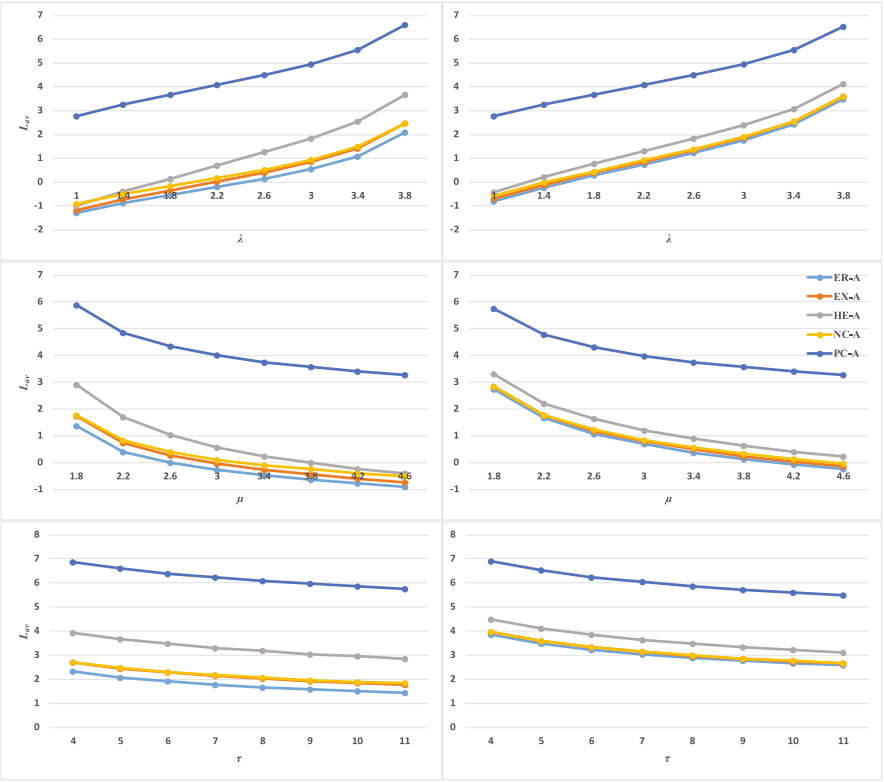

Fig. 4 $ln(L_{av})$ vs λ, μ, τ for ER-S (left side) and HE-S (right side)

6.1 The effect of parameters on performance measures

We investigate how the performance measures in (16)-(22) behave under various the arrival processes and service time distributions. For this purpose, the reorder point and the maximum inventory level are fixed by $s = 3$ and $S = 7$, respectively. The values of the other system parameters are varied as given in Table 1.

In the considered model growing the rate of customers (λ) leads to increasing the P_{lost} in case HE-S distribution for all arrival processes (see right side of Fig. 2). However, this phenomenon is not observed for ER-S distribution (see left side of Fig. 2), i.e. except PC-A arrival process, P_{lost} increases until the rate of customers reaches a certain value, after which it begins to decrease. Such behavior can be explained as follows: with increasing the rate of customers, the length of the queue reaches a threshold value and therefore the urgent replenishment will switch on, and thereby the level of inventory of the system increases, i.e. the P_{lost} is decreased. Growing the service rate of customers (μ) leads to increasing the P_{lost} in case ER-S for all arrival processes except ER-A (see left side of Fig. 2), i.e. for the ER-A process this measure increases at low values of the service rate, and after a certain value

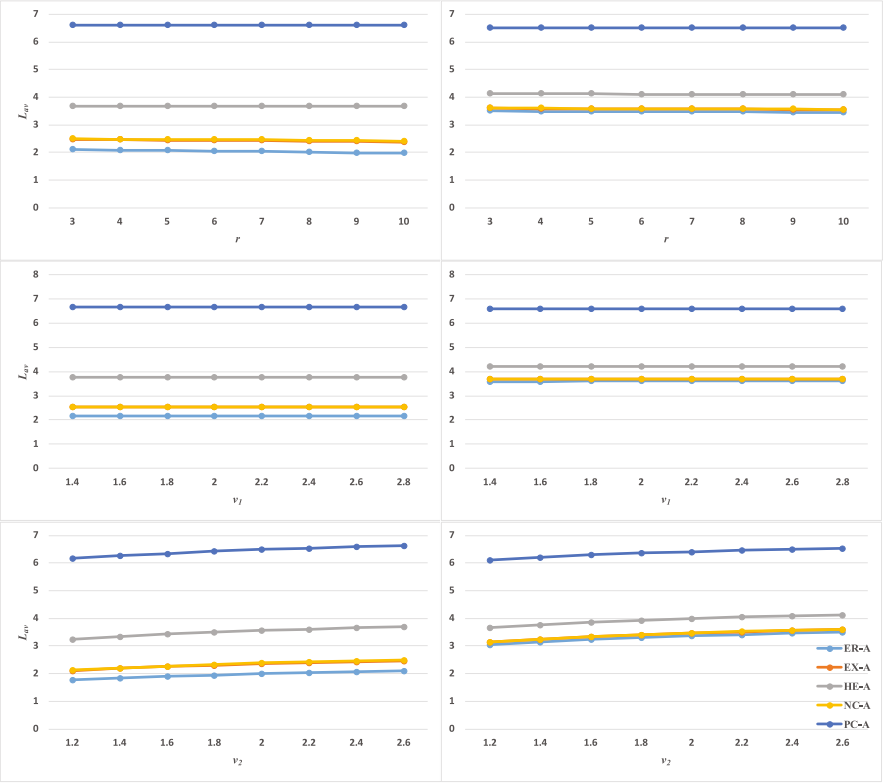

Fig. 5 $ln(L_{av})$ vs r, ν_1, ν_2 for ER-S (left side) and HE-S (right side)

of the service rate it begins to decrease. Such behavior of P_{lost} is unexpected. Similar behavior is observed in case HE-S for ER-A, EX-A and NC-A processes (see left side of Fig. 2). Growing the impatience rate of customers (τ) leads to decreasing the P_{lost} in case HE-S distribution for all arrival processes (see right side of Fig. 2). It was expected behavior of P_{lost} since leaving the queue of customers without items do not reduce the inventory level. In case ER-S distribution for PC-A, NC-A and EX-A arrival processes (see left side of Fig. 2), P_{lost} is increased at low values of the impatience rate, and after a certain value of the impatience rate it begins to decrease. Note that the values of P_{lost} are essentially high in case HE-S distribution.

Growing the threshold for switching urgent replenishment (r) leads to increasing the P_{lost} in both cases of HE-S and ER-S distributions for all arrival processes (see Fig. 3). The values of P_{lost} for various service distributions essentially differ each other. Growing the both rates of regular and urgent replenishment rates (ν_1, ν_2) leads to decreasing the P_{lost} in both HE-S and ER-S distributions for all arrival processes (see Fig. 3). The values of P_{lost} for various service distributions essentially differ each other.

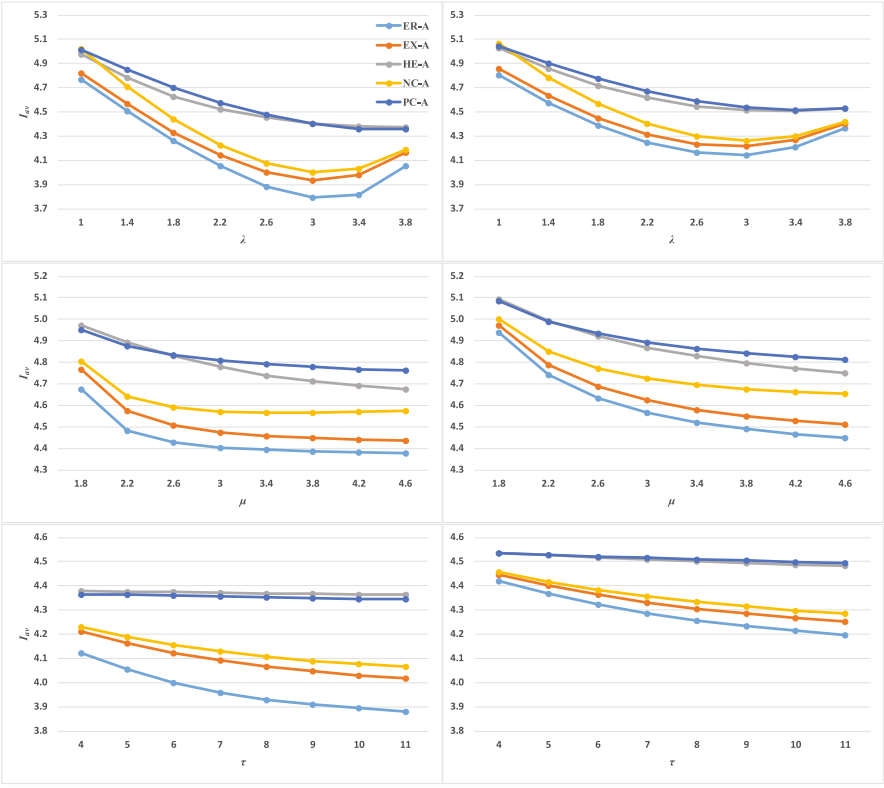

Fig. 6 I_{av} vs λ, μ, τ for ER-S (left side) and HE-S (right side)

The average number of customers in the system (L_{av}) is increasing function versus the rate of customers in both cases HE-S and ER-S distributions for all arrival processes; at the same time, this measure is decreasing function versus both service and impatience rates (see Fig. 4). Such behavior of this measure is expected. The average number of customers in the system (L_{av}) is almost constant versus the threshold for switching urgent replenishment and the rate of regular replenishment; however, it is slightly increased versus the rate of urgent replenishment (see Fig. 5).

At low values of the arrival rate of customers the average inventory level (I_{av}) is decreasing function and after its a certain value it begins to increase in both cases HE-S and ER-S distributions for all arrival processes (see Fig. 6). Such behavior of this measure was expected: with increasing the rate of customers, the length of the queue reaches a threshold value (r) to switch urgent replenishment, and thereby the level of inventory of the system increases. As it was expected, this measure is decreasing function versus both service and impatience rates (see Fig. 6). The average inventory level (I_{av}) is decreasing function versus the threshold for switching urgent replenishment in both cases HE-S and ER-S distributions

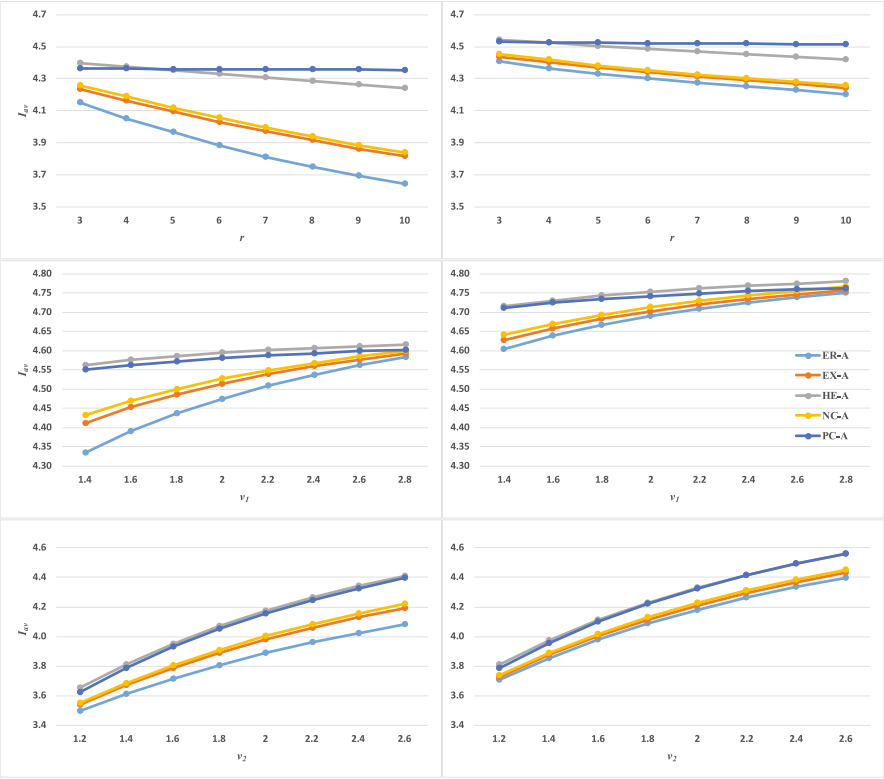

Fig. 7 I_{av} vs r, ν_1, ν_2 for ER-S (left side) and HE-S (right side)

for all arrival processes; at the same time, it is increasing function versus both rates of regular and urgent replenishments (see Fig. 7).

At low values of the arrival rate of customers the average volume of delivers via regular order (V_r) is increasing function and after its a certain value it begins to decrease in both cases HE-S and ER-S distributions for all arrival processes (see Fig. 8). Such behavior of this measure is explained as follows: with increasing the rate of customers, the length of the queue reaches a threshold value (r) to switch urgent replenishment (i.e. at this point regular order canceled), and thereby the average volume of delivers via regular order decreases. As it was expected, this measure is increasing function versus both service and impatience rates (see Fig. 8). The average volume of delivers via regular order (V_r) is increasing function versus parameter r in both cases HE-S and ER-S distributions for all arrival processes; this measure is decreasing function versus both service and impatience rates (see Fig. 9). Note that values of this measure in case ER-S distribution for ER-A, EX-A and NC-A arrivals are essentially high than in case HE-S distribution.

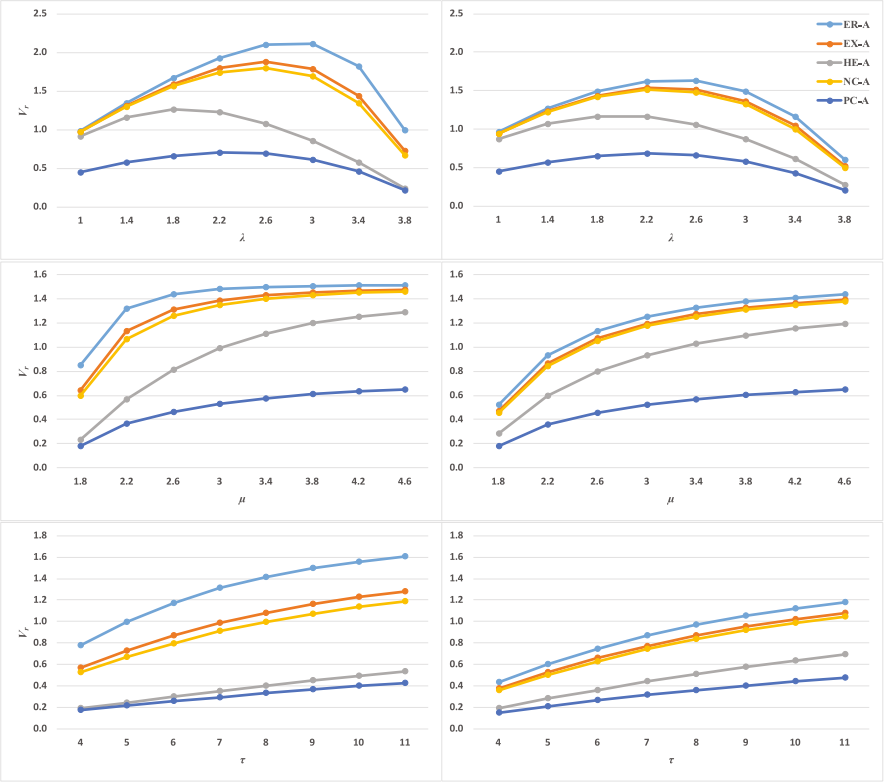

Fig. 8 V_r vs λ, μ, τ for ER-S (left side) and HE-S (right side)

The average volume of delivers via urgent order (V_u) is increasing function versus arrival rate of customers while it is decreasing function versus both service and impatient rates in both cases HE-S and ER-S distributions for all arrival processes (see Fig. 10). As it was expected, this measure is decreasing function versus parameter r in both cases HE-S and ER-S distributions for all considered arrival processes; it is almost constant versus rate of regular replenishments while it is decreasing function versus rate of urgent replenishments (see Fig. 11).

The average intensity of regular orders (RR_r) is increasing function versus arrival rate of customers for their low values and after its a certain value it begins to decrease in both cases HE-S and ER-S distributions for all arrival processes types (see Fig. 12). Such behavior of this measure is explained as follows: with increasing the rate of customers, the length of the queue reaches a threshold value (r) to switch urgent replenishment (i.e. at this point regular order canceled), and thereby the average intensity of regular orders decreases. As it was expected, this measure is increasing function versus both service and impatience rates (see Fig. 12). As it was expected, the average intensity of regular orders (RR_r) is increasing function versus parameter r in both cases HE-S and ER-S distributions

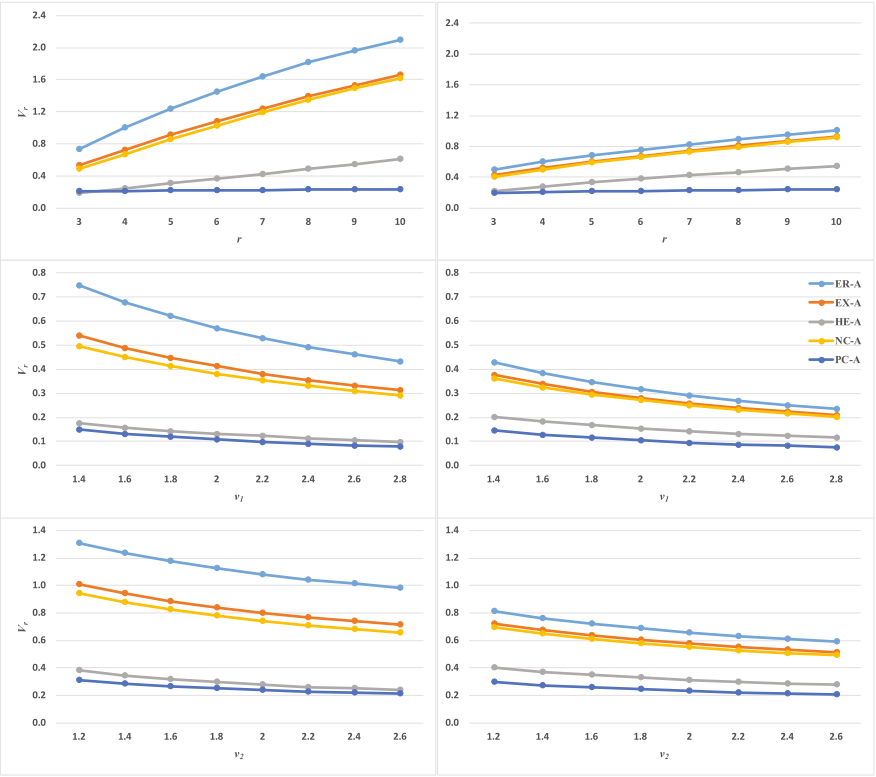

Fig. 9 V_r vs r, ν_1, ν_2 for ER-S (left side) and HE-S (right side)

for all arrival processes; this measure is almost linearly increasing function versus rate of regular replenishments while it is linearly decreasing function versus rate of urgent replenishments. Note that values of this measure in case ER-S distribution for ER-A, EX-A and NC-A arrivals are essentially high than in case HE-S distribution (see Fig. 13).

The average intensity of urgent orders (RR_u) is increasing function versus arrival rate of customers in both cases HE-S and ER-S distributions for all arrival processes; this measure is decreasing function versus both service and impatience rates because that with increasing these parameters chances of reaching value of length of queue to threshold (r) is decreased (see Fig. 14). The average intensity of urgent orders (RR_u) is decreasing function versus parameter r in both cases HE-S and ER-S distributions for all arrival processes; this measure is almost constant versus rate of regular replenishments while it is linearly increasing function versus rate of urgent replenishments (see Fig. 15).

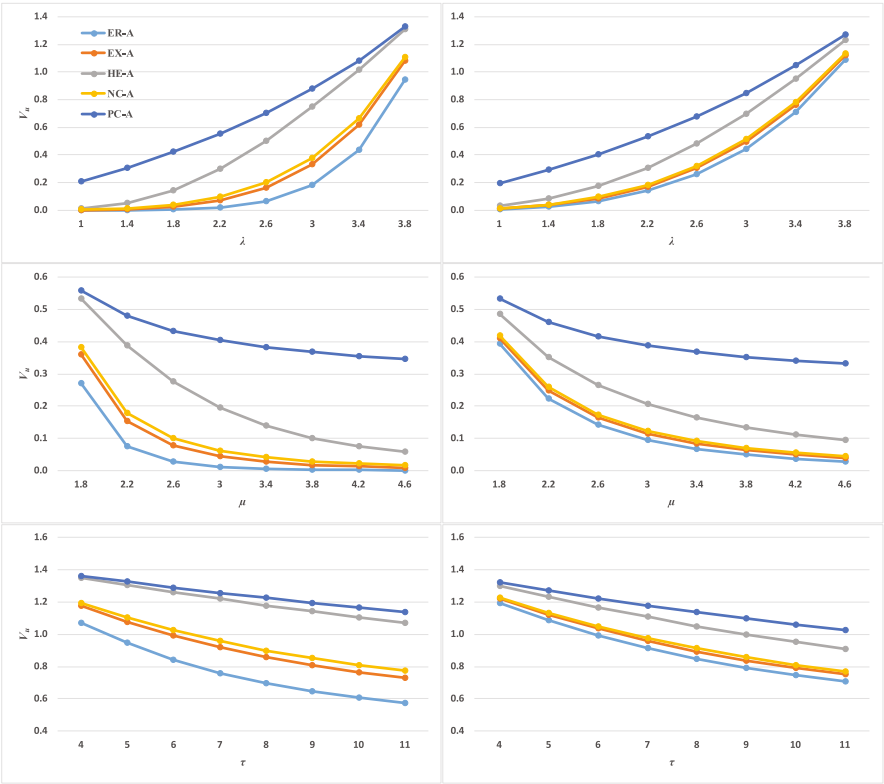

Fig. 10 V_u vs λ, μ, τ for ER-S (left side) and HE-S (right side)

6.2 Comparison of two systems: state dependent and state independent

In this section, we compare the queueing-inventory model with state-dependent replenishment policy described in Sect. 2 with the queueing-inventory model with state-independent replenishment policy in Sect. 5 in terms of some performance measures.

Tables 3 and 4 show the results for both systems. That is, the columns with **S-D** include the results for the system with the state-dependent replenishment policy and the columns with **S-I** include the results for the system with the state-independent replenishment policy. In order to compare the two systems, all parameter values except the replenishment rate are taken as the same for both systems. The values of the parameters are varied for the system with **S-D** as given in Table 1 and for the system with **S-I** as given in Table 2. In other words, in the system with **S-D**, the regular replenishment rate is fixed with $\nu_1 = 1$ and the urgent replenishment rate is fixed with $\nu_2 = 2.5$ depending on the thereshold value r, while in the system with **S-I**, since the urgent order is ignored, there is a single replenishment

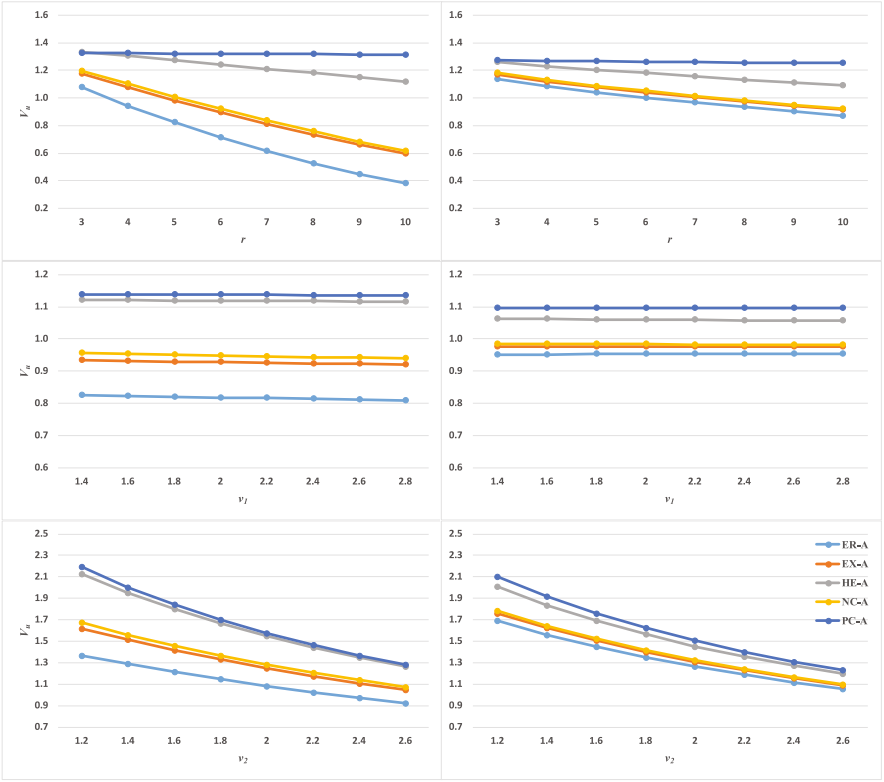

Fig. 11 V_u vs r, ν_1, ν_2 for ER-S (left side) and HE-S (right side)

rate and it is fixed with $\nu = 1$. The reorder point and the maximum inventory level are fixed by $s = 3$ and $S = 7$, respectively, for the both sysyems.

First of all, we would like to note that. In order to compare the two systems, we considered a single average volume of deliveries via order by adding the average volume calculated separately for the regular order and the urgent order in the system with **S-D**, $V = V_r + V_u$. Similarly, we considered a single average intensity by adding the average intensity calculated separately for the regular order and the urgent order in the system with **S-D**, $R_R = RR_r + RR_u$.

When we look at all cases in Tables 3 and 4, it is seen that the values of the probability of customers leaving the system are higher and the values of the average number of customers in the system are lower in the system with **S-I** compared to the system with **S-D**. The average number of items in the state-independent system is less compared to that in the state-dependent system. When the average volume of deliveries and the average intensity of the order of both systems are examined, it is found that the values of the average volume of deliveries are higher and the values of the average intensity of the order are lower in the system with **S-I**. In other words,

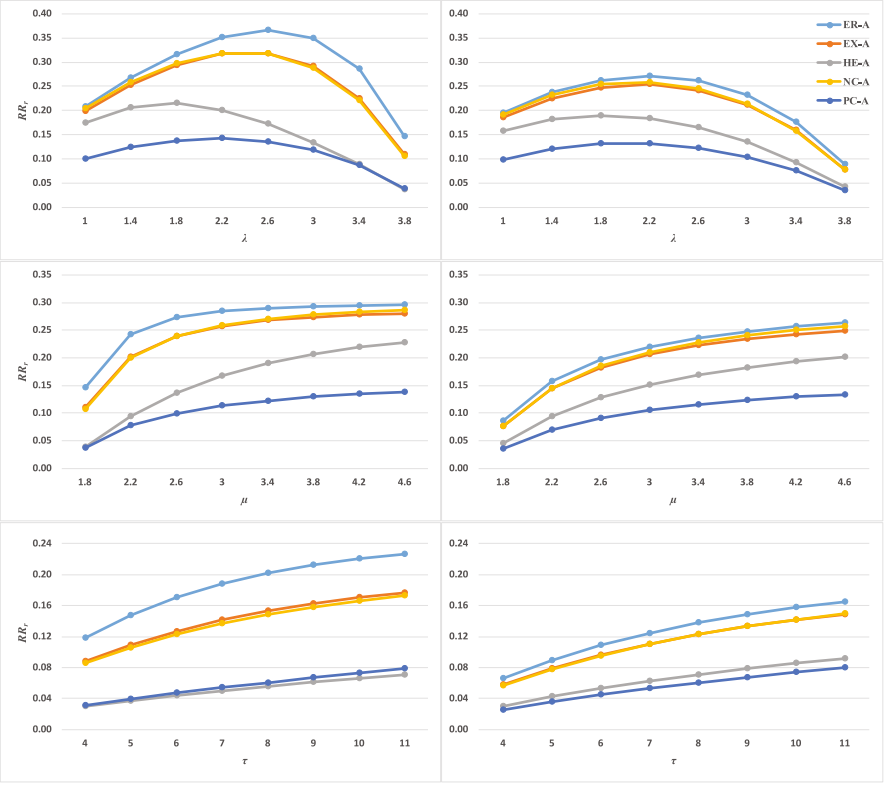

Fig. 12 RR_r vs λ, μ, τ for ER-S (left side) and HE-S (right side)

when the case of the urgent order is ignored (the system with **S-I**), larger orders are placed and the ordering frequency is lower.

6.3 Optimization

We define two functions for the expected total cost and discuss optimum inventory policies for some system parameters. That is, the expected total cost function for the system with **S-D** is given in (40) and the expected total cost function for the system with **S-I** is given in (41).

For the system with state-dependent (**S-D**), the function is given by

$$ETC_D = \left[k_r + c_r V_r\right] RR_r + \left[k_u + c_u V_u\right] RR_u + c_h I_{av} + c_l \tau P_{lost} + c_w L_{av}. \quad (40)$$

where k_r (k_u) is the fixed price of one regular (urgent) order, c_r (c_u) is the unit price of the regular (urgent) order, c_h is the holding price per item in the inventory per unit of time, c_l is the cost incured due to the loss of a customer and c_w is the waiting cost of a customer in the system.

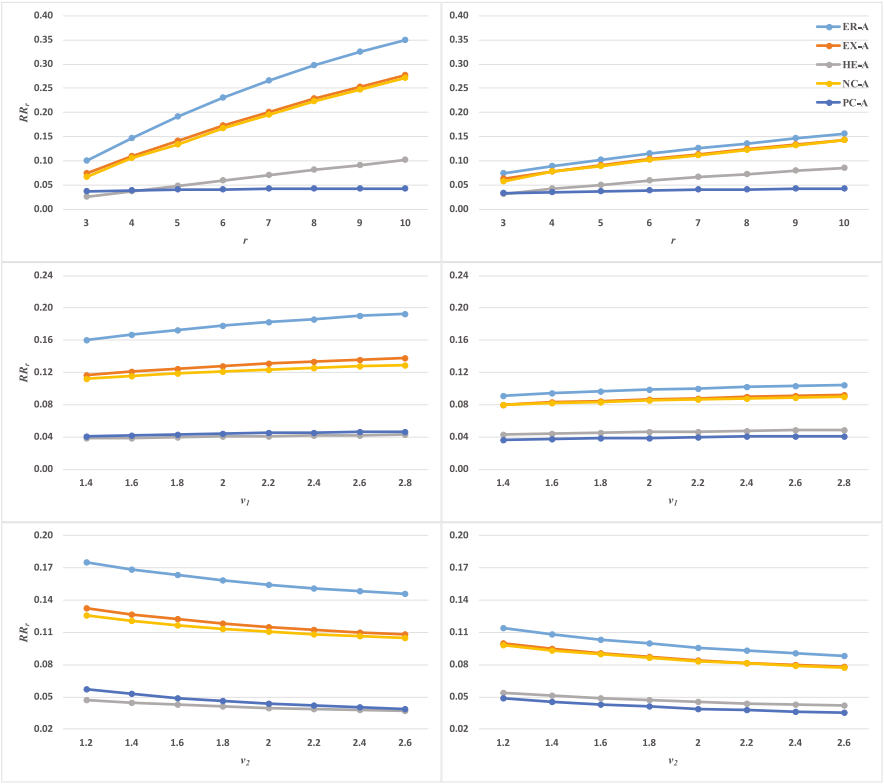

Fig. 13 RR_r vs r, v_1, v_2 for ER-S (left side) and HE-S (right side)

For the system with state-independent (**S-I**), the function is given by

$$ETC_I = \left[k + c\ \hat{V}\right]\hat{R}_R + c_h\hat{I}_{av} + c_l\tau\hat{P}_{lost} + c_w\hat{L}_{av}, \tag{41}$$

where k is the fixed price of one order and c is the unit price of the order.

To find the optimum values of the inventory level (that minimize ETC_D), we fix $\lambda = 2$, $\mu = 4$, $v_1 = 1$, $v_2 = 2.5$ and $r = 3$, and vary the reorder points $s = 3, 5, 7$ and the impatient rates $\tau = 1, 3, 6$. We fix also the unit values of the defined above costs by $k_r = 10$, $k_u = 30$, $c_r = 15$, $c_u = 45$, $c_h = 10$, $c_l = 150$ and $c_w = 80$.

To find the optimum values of the inventory level (that minimize ETC_I), we fix $v = 1$ and the unit values of the defined costs by $k = 10$, $c = 15$. Other parameter values $(\lambda, \mu, r, s, \tau)$ and cost values (c_h, c_l, c_w) are the same as those defined above.

Under various distributions of the service times and the inter-arrival times, the optimum values of the maximum inventory level S that minimize the expected total

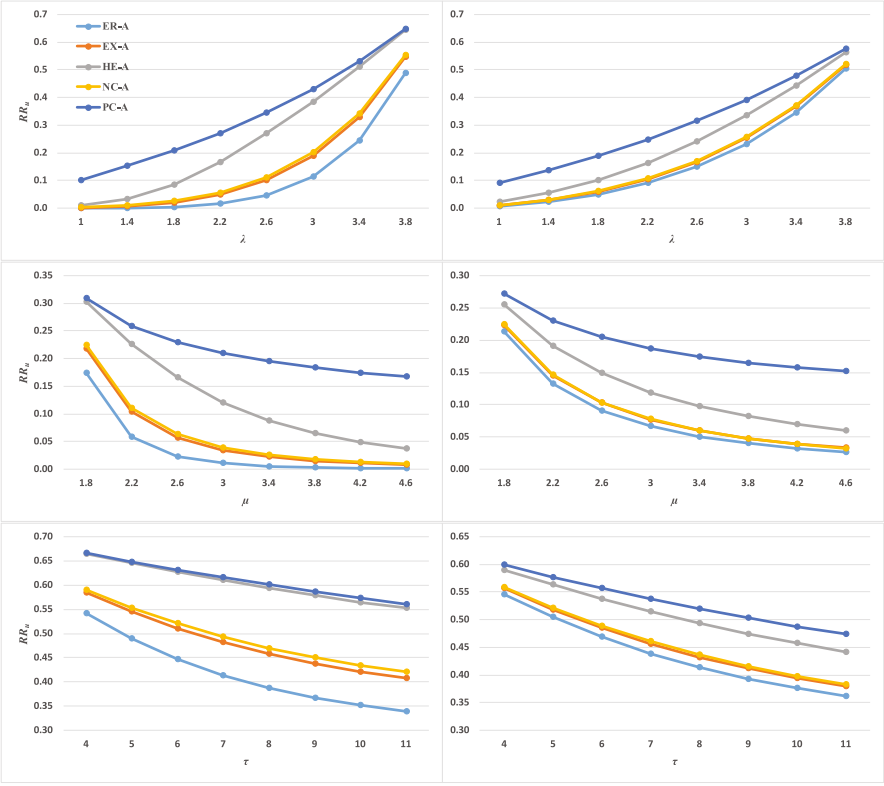

Fig. 14 RR_u vs λ, μ, τ for ER-S (left side) and HE-S (right side)

cost are given in Tables 5, 6 and 7. At this point, we would like to emphasize that the values given in parentheses are the results of the state-independent system (**S-I**).

From the Tables 5, 6 and 7, we conclude that for specific *MAP* (except PC-A process) the optimal solution S^* almost same for various service times distributions. Also note that in case PC-A process (see Tables 5, 6 and 7) and NC-A process (see Tables 6 and 7) for some initial data optimal solution does not exist, i.e. ETC_D (also ETC_I) is un-bounded increasing function (this is indicated by the symbol "-").

When the two systems are compared, it is seen in Tables 5, 6 and 7 that the obtained optimum values of S for the system with **S-I** (given in parentheses) are equal to or greater than the optimum values of S in the system with **S-D**. It can be said that the difference between the values of S is affected by the impatience rate τ, the variation in service times and the variation in interarrival times, and positively correlated arrivals. These differences are more pronounced in the case where the impatience rate is low, in the case of HE-S, in the case of HE-A, or in the case of PC-A. In other words, in these scenarios, the maximum inventory level is desired to be higher in the system with **S-I**. Finally, it is observed that the values of ETC_I

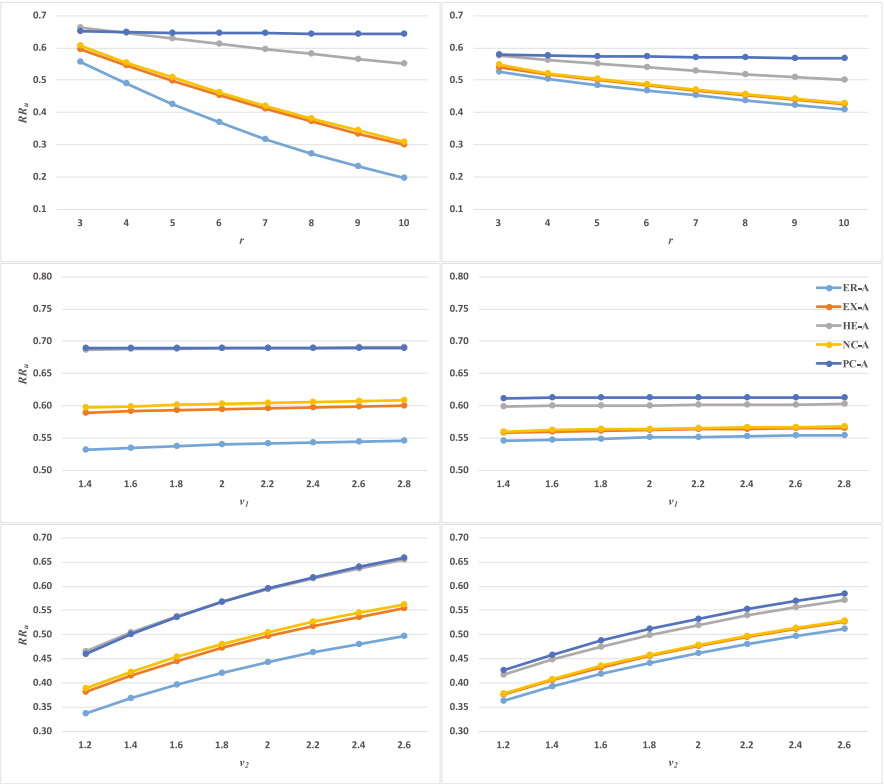

Fig. 15 RR_u vs r, ν_1, ν_2 for ER-S (left side) and HE-S (right side)

Table 2 The values of the system parameters for the case of state-independent

As It Is Varied	It Is Fixed
The arrival rate: λ	$r = 4, \mu = 4, \nu = 1, \tau = 5$
The service rate: μ	$r = 4, \lambda = 1.6, \nu = 1, \tau = 5$
The threshold point to switch order mode: r	$\lambda = 3.8, \mu = 4, \nu = 1, \tau = 5$

(given in parentheses) are greater than the values of ETC_D even when the values of S are equal.

7 Conclusion

An infinite QIS model with a queue-dependent replenishment policy is proposed, in which customers arrive according to the *MAP* and customer service times follow the *PH* distribution. Customers in the queue lose patience as inventory levels drop

Table 3 Comparison of the two systems for the performance measures under arrival ER-A

		S-D	S-I	S-D	S-I	S-D	S-I	S-D	S-I	S-D	S-I
PH	λ	P_{lost}	$\hat{P}_{lost}$	L_{av}	$\hat{L}_{av}$	I_{av}	$\hat{I}_{av}$	V	$\hat{V}$	R_R	$\hat{R}_R$
ER-S	2.2	0.044	0.046	0.807	0.805	4.054	4.039	1.950	1.970	0.368	0.364
	3	0.087	0.107	1.709	1.651	3.791	3.660	2.293	2.465	0.464	0.435
	3.8	0.088	0.192	8.058	5.454	4.054	3.368	1.943	2.838	0.636	0.486
HE-S	2.2	0.046	0.061	2.127	2.032	4.249	4.137	1.760	1.897	0.362	0.342
	3	0.082	0.131	5.857	5.054	4.147	3.805	1.927	2.347	0.463	0.403
	3.8	0.097	0.219	32.806	18.595	4.366	3.539	1.686	2.706	0.594	0.449
PH	μ	P_{lost}	$\hat{P}_{lost}$	L_{av}	$\hat{L}_{av}$	I_{av}	$\hat{I}_{av}$	V	$\hat{V}$	R_R	$\hat{R}_R$
ER-S	1.8	0.014	0.030	3.959	3.046	4.676	4.439	1.122	1.451	0.320	0.285
	2.6	0.019	0.021	0.992	0.981	4.426	4.405	1.465	1.495	0.296	0.292
	3.4	0.018	0.018	0.617	0.616	4.393	4.389	1.503	1.508	0.295	0.294
HE-S	1.8	0.020	0.050	15.562	9.160	4.940	4.610	0.914	1.349	0.299	0.251
	2.6	0.022	0.035	2.958	2.681	4.632	4.514	1.277	1.426	0.287	0.268
	3.4	0.021	0.027	1.467	1.421	4.521	4.465	1.394	1.463	0.287	0.278
PH	r	P_{lost}	$\hat{P}_{lost}$	L_{av}	$\hat{L}_{av}$	I_{av}	$\hat{I}_{av}$	V	$\hat{V}$	R_R	$\hat{R}_R$
ER-S	3	0.073		8.128		4.153		1.814		0.657	
	5	0.101	0.192	7.953	5.454	3.965	3.368	2.059	2.838	0.617	0.486
	7	0.124		7.666		3.814		2.257		0.584	
HE-S	3	0.092		33.048		4.406		1.636		0.601	
	5	0.102	0.219	32.587	18.595	4.333	3.539	1.726	2.706	0.588	0.449
	7	0.110		32.176		4.277		1.794		0.579	

to zero. The replenishment process is governed by the (s, S)-policy as follows: if the inventory level drops to the reorder point and the number of customers in the system is less than a predetermined threshold, then a regular order is made; when the inventory level drops to the reorder point and the number of customers in the system is greater than or equal to the specified threshold, the regular order is immediately cancelled and an urgent order is sent. Lead times follow exponential distributions with different averages depending on the order type. An easily checkable condition for the stability of the constructed multi-dimensional Markov chain is established and its probabilistic meaning is explained. Steady state probabilities were obtained using a matrix geometric method, and key performance indicators were calculated using these probabilities.

The queue-independent replenishment policy is also considered for the studied queueing-inventory model. So, we ignore the threshold value depend on the number of customers in the system (Namely, the urgent order situation is ignored). Regardless of the number of customers in the system, an order is placed when the inventory number drops to s. Except for the replenishment policy, all assumptions of the studied model are valid for the state-independent queueing-inventory model.

Table 4 Comparison of the two systems for the performance measures under arrival HE-A

		S-D	S-I	S-D	S-I	S-D	S-I	S-D	S-I	S-D	S-I
PH	λ	P_{lost}	$\hat{P}_{lost}$	L_{av}	$\hat{L}_{av}$	I_{av}	$\hat{I}_{av}$	V	$\hat{V}$	R_R	$\hat{R}_R$
ER-S	2.2	0.047	0.080	2.005	1.891	4.525	4.304	1.524	1.802	0.369	0.320
	3	0.056	0.141	6.325	5.448	4.407	3.873	1.605	2.296	0.520	0.398
	3.8	0.058	0.209	39.607	22.830	4.377	3.460	1.551	2.755	0.683	0.470
HE-S	2.2	0.054	0.088	3.669	3.372	4.616	4.375	1.469	1.759	0.348	0.305
	3	0.075	0.155	10.786	8.960	4.512	3.984	1.575	2.225	0.471	0.374
	3.8	0.088	0.230	61.286	34.926	4.527	3.614	1.513	2.650	0.606	0.436
PH	μ	P_{lost}	$\hat{P}_{lost}$	L_{av}	$\hat{L}_{av}$	I_{av}	$\hat{I}_{av}$	V	$\hat{V}$	R_R	$\hat{R}_R$
ER-S	1.8	0.008	0.045	18.482	10.147	4.972	4.546	0.764	1.374	0.342	0.266
	2.6	0.020	0.044	2.819	2.420	4.830	4.610	1.087	1.380	0.302	0.258
	3.4	0.028	0.042	1.278	1.215	4.739	4.628	1.252	1.390	0.279	0.256
HE-S	1.8	0.020	0.062	27.655	14.660	5.093	4.720	0.773	1.291	0.300	0.235
	2.6	0.028	0.054	5.074	4.277	4.922	4.709	1.061	1.331	0.277	0.240
	3.4	0.032	0.049	2.422	2.251	4.829	4.693	1.193	1.356	0.267	0.244
PH	r	P_{lost}	$\hat{P}_{lost}$	L_{av}	$\hat{L}_{av}$	I_{av}	$\hat{I}_{av}$	V	$\hat{V}$	R_R	$\hat{R}_R$
ER-S	3	0.055		39.645		4.395		1.519		0.687	
	5	0.061	0.209	39.553	22.830	4.352	3.460	1.581	2.755	0.678	0.470
	7	0.068		39.403		4.306		1.640		0.667	
HE-S	3	0.085		61.436		4.545		1.484		0.609	
	5	0.091	0.230	61.125	34.926	4.506	3.614	1.538	2.650	0.602	0.436
	7	0.096		60.801		4.469		1.583		0.595	

Under various arrival processes and service time distributions, the behavior of the some performance measures and optimum inventory policy are discussed. Numerical experiments demonstrate the influence of positive and negative correlations on the system performance measures, and show that the variability in the servis times and the inter-arrival times play a key role in the behavior of system performance measures. In addition, the expected total cost calculated based on the system performance metrics is also affected from the variability and correlations. When the two systems are compared, it is seen that the values of the probability of customers leaving the system are higher and the values of the average number of customers in the system are lower in the state-independent system with compared to the state-dependent system. Also, in the comparative studies, it is seen that the obtained optimum values of S for the state-independent system are equal to or greater than the optimum values of S in the state-dependent system. It can be said that the difference between the values of S is affected by the the variability in service times and the variability in interarrival times, and positively correlated arrivals. These differences are more pronounced in the case of high variability and in the case of a positive correlation.

Table 5 Optimum values of S^* and ETC^* for reorder point $s = 3$

		$\tau = 6$		$\tau = 3$		$\tau = 1$	
MAP	**PH**	S^*	ETC^*	S^*	ETC^*	S^*	ETC^*
ER-A	ER-S	7	132.561	7	131.227	7	128.859
		(8)	(134.962)	(8)	(135.392)	(9)	(140.419)
	EX-S	7	140.739	7	139.634	7	137.766
		(7)	(144.067)	(8)	(144.967)	(9)	(150.909)
	HE-S	6	214.011	7	217.027	7	220.741
		(6)	(216.997)	(8)	(227.313)	(10)	(245.967)
EX-A	ER-S	7	147.040	7	145.757	7	144.375
		(8)	(152.706)	(8)	(154.620)	(9)	(162.612)
	EX-S	7	156.346	7	155.398	7	154.547
		(8)	(162.405)	(8)	(164.946)	(10)	(174.136)
	HE-S	6	230.469	7	234.812	8	239.959
		(6)	(233.685)	(8)	(247.471)	(11)	(269.626)
HE-A	ER-S	7	205.765	8	209.602	8	214.459
		(8)	(216.007)	(10)	(233.941)	(13)	(260.413)
	EX-S	7	222.082	8	227.160	8	233.045
		(7)	(230.691)	(10)	(251.139)	(14)	(279.410)
	HE-S	5	311.840	7	330.089	9	344.328
		(5)	(308.748)	(9)	(353.803)	(15)	(397.009)
NC-A	ER-S	8	158.953	8	157.583	7	156.180
		(8)	(165.479)	(8)	(167.447)	(9)	(175.698)
	EX-S	8	167.464	8	166.384	7	165.489
		(8)	(173.918)	(8)	(176.541)	(10)	(185.874)
	HE-S	6	238.665	6	243.667	8	248.391
		(6)	(242.204)	(8)	(256.311)	(11)	(278.992)
PC-A	ER-S	–	–	14	4128.326	22	4205.208
		(–)	(–)	(29)	(4285.574)	(51)	(4491.633)
	EX-S	–	–	15	4146.899	23	4231.566
		(–)	(–)	(30)	(4299.686)	(52)	(4510.130)
	HE-S	–	–	17	4250.093	28	4355.975
		(–)	(–)	(31)	(4397.769)	(55)	(4627.718)

Table 6 Optimum values of S^* and ETC^* for reorder point $s = 5$

		$\tau = 6$		$\tau = 3$		$\tau = 1$	
MAP	**PH**	S^*	ETC^*	S^*	ETC^*	S^*	ETC^*
ER-A	ER-S	8	137.011	8	136.294	8	135.012
		(8)	(138.845)	(8)	(139.099)	(8)	(141.854)
	EX-S	8	145.496	8	144.907	8	143.884
		(8)	(148.005)	(8)	(148.474)	(9)	(151.787)
	HE-S	7	222.214	7	223.792	8	225.495
		(7)	(224.627)	(8)	(231.200)	(10)	(243.320)
EX-A	ER-S	8	151.423	8	150.741	8	149.958
		(8)	(155.686)	(8)	(156.853)	(9)	(161.285)
	EX-S	8	161.139	8	160.609	8	160.053
		(8)	(165.538)	(8)	(167.088)	(9)	(172.241)
	HE-S	7	238.887	7	241.234	8	243.705
		(7)	(241.122)	(8)	(250.395)	(10)	(265.153)
HE-A	ER-S	8	211.506	8	213.289	8	215.768
		(8)	(217.662)	(9)	(230.427)	(12)	(248.705)
	EX-S	8	228.653	8	230.958	8	233.976
		(8)	(233.444)	(9)	(247.911)	(12)	(267.634)
	HE-S	–	–	8	334.949	9	343.254
		(–)	(–)	(9)	(352.789)	(13)	(385.223)
NC-A	ER-S	8	163.131	8	162.425	8	161.745
		(8)	(168.022)	(8)	(169.230)	(9)	(174.560)
	EX-S	8	171.799	8	171.237	8	170.759
		(8)	(176.612)	(8)	(178.218)	(9)	(184.335)
	HE-S	–	–	8	249.735	8	251.951
		(–)	(–)	(8)	(258.774)	(10)	(273.798)
PC-A	ER-S	–	–	11	4103.099	16	4144.778
		(–)	(–)	(25)	(4242.615)	(41)	(4403.350)
	EX-S	–	–	12	4120.599	17	4170.076
		(–)	(–)	(25)	(4258.278)	(43)	(4424.532)
	HE-S	–	–	14	4224.578	22	4296.760
		(–)	(–)	(28)	(4363.490)	(47)	(4554.197)

Table 7 Optimum values of S^* and ETC^* for reorder point $s = 7$

		$\tau = 6$		$\tau = 3$		$\tau = 1$	
MAP	**PH**	S^*	ETC^*	S^*	ETC^*	S^*	ETC^*
ER-A	ER-S	9	149.334	9	148.964	9	148.294
		(9)	(150.630)	(9)	(150.735)	(9)	(151.947)
	EX-S	9	157.915	9	157.620	9	157.098
		(9)	(159.602)	(9)	(159.810)	(9)	(161.302)
	HE-S	9	236.365	9	236.890	9	237.512
		(9)	(237.719)	(9)	(241.436)	(10)	(248.464)
EX-A	ER-S	9	163.651	9	163.318	9	162.933
		(9)	(166.280)	(9)	(166.869)	(10)	(169.279)
	EX-S	9	173.478	9	173.222	9	172.948
		(9)	(176.071)	(9)	(176.863)	(10)	(179.730)
	HE-S	9	253.157	9	253.994	9	255.020
		(–)	(–)	(9)	(259.456)	(10)	(268.672)
HE-A	ER-S	9	224.310	9	225.226	9	226.456
		(9)	(225.653)	(10)	(234.156)	(11)	(246.541)
	EX-S	9	241.552	9	242.713	9	244.192
		(9)	(242.016)	(10)	(251.736)	(11)	(265.244)
	HE-S	–	–	9	345.897	9	350.496
		(–)	(–)	(9)	(357.387)	(12)	(381.228)
NC-A	ER-S	10	176.944	–	–	10	176.392
		(–)	(–)	(10)	(179.580)	(10)	(181.504)
	EX-S	–	–	–	–	–	–
		(–)	(–)	(–)	(–)	(10)	(190.703)
	HE-S	–	–	–	–	–	–
		(–)	(–)	(–)	(–)	(10)	(276.940)
PC-A	ER-S	–	–	11	4100.314	13	4121.213
		(–)	(–)	(22)	(4213.629)	(35)	(4338.899)
	EX-S	–	–	12	4115.253	14	4142.575
		(–)	(–)	(22)	(4229.907)	(36)	(4361.134)
	HE-S	–	–	13	4214.397	18	4262.734
		(–)	(–)	(25)	(4338.525)	(41)	(4496.716)

As a future research direction, one can point out the study of a single-source QIS system with a hybrid queue-dependent replenishment policy, i.e. if the queue length is less than some threshold, then the (s, Q)-policy is used, $Q = S - s$; if the queue length is greater than or equal to this threshold, then the (s, Q)-policy is cancelled and the (s, S)-policy is used with the same lead time. Another extension of this work could be to consider the QIS model with a batch Markov process (*BMAP*) and with (without) perishable inventories. The method proposed here has a limitation in that it can be applied to QIS models with an infinite queue. Studying models with a finite queue requires further research. This task is also the subject of our further research.

Funding The authors declare that this paper is unfunded.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

References

Baek JW (2024) On the control policy of a queuing-inventory system with variable inventory replenishment speed. Mathematics 2024(12):194. https://doi.org/10.3390/math12020194
Berman O, Kim E (1999) Stochastic models for inventory management at service facilities. Commun Stat Stoch Mod 15(4):695–718
Berman O, Sapna KP (2000) Inventory management at service facilities for systems with arbitrary distributed service times. Commun Stat Stoch Mod 16(3):343–360
Bijvank M, Iris FA, Vis I (2011) Lost-sales inventory theory: a review. Eur J Oper Res 215:1–13
Chakravarthy SR (2022a) Introduction to matrix-analytic methods in queues 1- Analytical and simulation approach- Basics. Wiley, London
Chakravarthy SR (2022b) Introduction to matrix-analytic methods in queues 2–Analytical and simulation approach–Queues and simulation. Wiley, London
Chakravarthy SR, Melikov A (2024) A new admission control scheme in queueing-inventory system with two priority classes of demands. Opsearch. https://doi.org/10.1007/s12597-024-00877-8
Dudin AN, Klimenok VI, Vishnevsky VM (2020) The theory of queueing systems with correlated flows. Springer, Basel, Switzerland
He QM (2014) Fundamentals of matrix-analytic methods. Springer, New York
He Q, Jewkes EM (2000) Performance measure of a make-to-order inventory-production system. IIE Trans 32:409–419
He Q, Jewkes EM, Buzacott J (2002) The value of information used in inventory control of a make-to-order inventory-production system. IIE Trans 34:999–1013
He Q, Jewkes EM, Buzacott J (2002) Optimal and near-optimal inventory control policies for a make-to-order inventory-production system. Eur J Prod Res 141:113–132
Jose KP, Nair SS (2017) Analysis of two production inventory systems with buffer, retrials and different production rates. J Indus Eng Int 13:369–380
Kim E (2005) Optimal inventory replenishment policy for a queueing system with finite waiting room capacity. Eur J Oper Res 161:256–274
Krishnamoorthy A, Lakshmy B, Manikandan R (2011) A survey on inventory models with positive service time. Opsearch 48:153–169
Krishnamoorthy A, Shajin D, Narayanan W (2021) Inventory with positive service time: a survey. In: Anisimov, V, Limnios N (eds) Advanced trends in queueing theory; Series of books mathematics and statistics sciences. V.2; ISTE & Wiley: London, pp 201-238
Latouche G, Ramaswami V (1999) Introduction to matrix analytic methods in stochastic modeling. SIAM, Philadelphia
Melikov A, Fatalieva M (1998) Situational inventory in counter-stream serving systems. Eng Simul 15:839–848
Melikov A, Molchanov A (1992) Stock optimization in transport/storage systems. Cybernetics 28(3):484–487
Melikov A, Ponomarenlo L, Aliyev I (2018) Markov models of systems with demands of two types and different restocking policies. Cybernet Syst Anal 54(6):900–917
Melikov A, Ponomarenlo L, Aliyev I (2019) Markov models of queuing-inventory systems with different types of retrial customers. J Autom Inf Sci 51(8):1–15
Melikov A, Mirzayev R, Nair SS (2022a) Numerical study of a queuing-inventory system with two supply sources and destructive customers. J Comput Syst Sci Int 61:581–598
Melikov A, Mirzayev R, Nair SS (2022b) Double sources queuing-inventory system with hybrid replenishment policy. Mathematics 10:2423
Melikov A, Mirzayev RR, Sztrik J (2023) Double sources QIS with finite waiting room and destructible stocks. Mathematics 11:226

Mine H, Osaki S (1970) Markovian decision processes. American Elsevier Publishing Company Inc, New York
Minner S (2003) Multiple-supplier inventory models in supply chain management: a review. Int J Production Econ 81–82:265–279
Neuts MF (1981) Matrix-geometric solutions in stochastic models: An algorithmic approach. John Hopkins University Press, Baltimore
Otten S, Daduna H (2022) Stability of queueing-inventory systems with customers of different priorities. Ann Oper Res 331(2):963–983
Rasmi K, Jacob MJ, Rumyantsev AS, Krishnamoorthy A (2021) A multi-server heterogeneous queuing-inventory system with class-dependent inventory access. Mathematics 9:1037
Rejitha KR, Jose KP (2017) A queuing-inventory system with *MAP*, retrials and different replenishment rates. Int J Pure Appl Math 117(11):289–297
Salini K, Arya PS, Manikandan R (2023) Queuing-inventory systems: Review. arXiv:2308.06518v3 [math. PR] 19 Sep 2023
Shajin D, Dudin AN, Dudina OS, Krishnamoorthy A (2020) A two-priority single server retrial queue with additional items. J Indus Manag Optim 16(6):2891–2912
Sigman K, Simchi-Levi D (1992) Light traffic heuristic for an $M/G/1$ queue with limited inventory. Ann Oper Res 40:371–380
Sugapriya C, Nithya M, Jeganathan K, Anbazhagan N, Joshi GP, Yang E, Seo S (2022) Analysis of stock-dependent arrival process in a retrial stochastic inventory system with server vacation. Processes 10(1):176
Sugapriya C, Nithya M, Jeganathan K, Selvakumar S, Harikrishnan T (2023) A comparative analysis of (s, Q) and (s, S) ordering policies in a queuing-inventory system with stock-dependent arrival and queue-dependent service process. Oper Res Decis 33(2):121–153. https://doi.org/10.37190/ord230207
Varghese DT, Shajin D (2018) State dependent admission of demands in a finite storage system. Int J Pure Appl Math 118(20):917–922