UNIVERSITÄT HAMBURG

M/M/1-systems with different inventory management policies and lost sales

M. K. Schwarz, C. Sauer, H. Daduna,
R. Kulik, R. Szekli

Preprint–No. 2003-07 September 2003

FACHBEREICH MATHEMATIK

SCHWERPUNKT MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

M/M/1-systems with different inventory management policies and lost sales

Maike Schwarz, Cornelia Sauer, Hans Daduna
University of Hamburg, Department of Mathematics
Center of Mathematical Statistics and Stochastic Processes,
Bundesstrasse 55, 20146 Hamburg, Germany

Rafał Kulik, Ryszard Szekli *
University of Wrocław, Mathematical Institute,
Pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland

September 2003

Abstract

We derive stationary distributions of joint queue length and inventory processes in explicit product form for various M/M/1-systems with inventory management with continuous review, Poisson demand, exponentially distributed lead times, and lost sales. These distributions are then used to calculate performance measures of the respective systems.

MSC 2000 Subject Classification: 60K25, 90B22, 90B05
Key words: queueing systems; inventory systems; performance analysis; inventory policy; queue lengths; service level; inventory level; stationary distribution; convex ordering

1 Introduction

The importance of inventory management for the quality of service (QoS) of today's service systems is generally accepted, and from the point of view of inventory availability there are many different classical definitions of quality ([Ter94] p. 232). On the other hand only recently papers contributed to the problem of how the classical performance measures (e.g. queue length, waiting time, etc.) are influenced by the management of attached inventory. An early contribution is [SSL92] where approximation procedures are used to find performance descriptions for models, in which the interaction of queueing for service and inventory control is incorporated. In a sequence of papers Berman and his coauthors ([BK99], [BK01], [BS00], [BS02]) investigate the behaviour of service systems with an attached inventory. Their approach can be characterized as follows. Define a Markovian system process and then use classical optimization methods to find the optimal control strategy of the inventory. All these models assume that the demand, which arrives during the time the inventory is depleted, is backordered. The models vary with respect to the lead time distribution, the service time distribution, waiting room size, order size and reorder policy.
All the work cited up to now is of continuous review structure for the inventory. The growing use of computer systems to control inventories has boosted the interest in continuous-review inventory

*All authors were supported by DAAD/KBN grant number D/02/32206

models [MP98]. The lost sales situation arises in many retail establishments [CS99], where the intense competition allows customers to choose another brand or to go to another store. However, the area of continuous review inventory systems with lost sales remains largely unexplored [MP98]. The literature on lost sales models is reviewed in the article of Mohebbi and Posner [MP98].

The aim of our research is to find explicit performance measures for a service system with an attached inventory under continuous review and lost sales as well as availability measures and service grades for the inventory. We model and analyse several single server queueing systems of M/M/1-type with an attached inventory. Customers arrive according to a Poisson process with intensity λ and each customer, who is served, needs exactly one item from the inventory and has an exponentially distributed service time with parameter μ. Consequently, the demand rate of the inventory is equal to λ if there are no customers waiting in queue otherwise the demand rate is equal to the service rate μ. The variable replenishment lead time, which is the time span between ordering of materials and receipt of the goods, is exponentially distributed with parameter ν. The entire order is received into stock at the same time. The type of inventory system is defined to be a continuous review system where the inventory state is inspected after every single demand event and orders are placed every time the inventory on hand reaches a reorder point r. The on-hand stock is the stock that is physically on the shelf. The systems under investigation differ with respect to the size of replenishment orders and the reorder policy. Each system has the limiting property that no customers are allowed to join the queue as long as the inventory is empty. This corresponds to the lost sales case of inventory management. However, customers are still admitted to the waiting room even if the number of customers in the system exceeds the inventory on hand.

The rest of the paper is organized as follows. Sections 2, 3 and 4 are devoted to infinite waiting room systems. In Section 2 we investigate a system with reorder point 0 and arbitrary random size of replenishment orders. In Sections 3 and 4 we investigate the analogue of the celebrated (r, Q)- and (r, S)-policies of inventory management in the context of service stations with attached inventory. In Sections 6.1, 6.2 and 6.3 the systems from Sections 2, 3 and 4 are investigated under the restriction of finite waiting rooms. It is important that we assume throughout the paper that during the time a replenishment order is outstanding the service place can be used as a waiting place by the customers in the system.
For each system we compute the steady state probability distribution and calculate the most important performance measures. It turnes out that a special feature of the above models is that the steady state probabilities are of product form. In case of infinite waiting room the limiting distributions of the queue length processes coincide in addition with that of the M/M/1/∞-system. This shows a very important invariance property for the queueing systems with inventory management and lost sales investigated in this paper.
Throughout the paper we will assume that unless otherwise specified an underlying probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is given where all random variables are defined on.

2 M/M/1/∞-system with inventory management, reorder point 0 and arbitrary size of replenishment orders

In the following definition we introduce our M/M/1/∞-system with inventory management, reorder point 0 and arbitrary size of replenishment orders.

Definition 2.1 *An M/M/1/∞-system with inventory management, reorder point* 0 *and arbitrary size of replenishment orders is a single server with infinite waiting room under FCFS regime. To the server there is a Poisson-λ-arrival stream, $\lambda \geq 0$. Customers request an amount of service time*

which is exponentially distributed with mean 1. Service is provided with intensity $\mu > 0$. All service and interarrival times are assumed to be independent.
The server needs for each customer exactly one item from an attached inventory. The on-hand inventory decreases by one at the moment of service completion. If the inventory is depleted after the service of a customer is completed, a replenishment order is instantaneously triggered. The replenishment lead time is exponentially distributed with parameter $\nu > 0$. After a replenishment order arrives the new inventory position is determined according to a discrete probability density function p on the positive integers, i.e. the size of a replenishment order is k with probability p_k, where $\sum_{k=1}^{\infty} p_k = 1$. The corresponding discrete distribution function and its tail distribution function will be denoted by F_p and $\overline{F}_p := 1 - F_p$, respectively. We abbreviate the probability that the size of a replenishment order is at least k units by q_k, i.e. $q_k = \overline{F}_p(k) = \sum_{h=k}^{\infty} p_h$. The mean order size is denoted by $\overline{p}$, i.e. $\overline{p} = \sum_{k=1}^{\infty} kp_k$.
During the time the server waits for a replenishment order to arrive, no customers are admitted to join the queue ("lost sales").
Let $X(t)$ denote the number of customers present at the server at time $t \geq 0$, either waiting or in service (queue length) and let $Y(t)$ denote the on-hand inventory at time $t \geq 0$. Then $Z = ((X(t), Y(t)), t \geq 0)$ is a continuous-time Markov process for the M/M/1/∞-system with inventory management, reorder point 0 and arbitrary size of replenishment orders. The state space of Z is $E_Z = \{(n, k) : n \in \mathbb{N}_0, k \in \mathbb{N}_0\}$.

The system described above generalizes the lost sales case of classical inventory management where customer demand is not backordered but lost in case there is no inventory on hand (see Tersine [Ter94] p. 207).

We consider the system of Definition 2.1 to be fundamental. It corresponds to the popular (r, S)- and (r, Q)-policies with $r = 0$ and random S, resp. Q. As was shown by [BK99], in case of exponential interarrival and service times and zero lead times $r > 0$ is suboptimal if customer demand is backordered and inventory holding costs are involved. An optimal order policy for that system only places an order when the inventory level drops to zero and the number of customers in the system exceeds some threshold value. The situation for stochastic lead times is different. Nevertheless, we will carry out in detail all the calculations for the system with arbitrary order size and reorder point 0 and show how to define the important measures of system performance. An outcome of this investigation will be that all except for one performance measure only depend on the mean order size $\overline{p}$ and not on the whole distribution F_p. Only the stationary mean inventory position depends on the second moment of F_p and will be shown to be minimal for fixed order size in Section 2.3.
Coming back to our earlier considerations it can be profitable to already trigger an order if there is still some inventory on stock in case of stochastic lead times if customer waiting costs are incurred. Because of our insights into the M/M/1/∞-system with inventory management, reorder point 0 and arbitrary size of replenishment orders it is justifiable to extend the system to early replenishments with reorder level $r \geq 0$ without admitting for variable order sizes. This results in the M/M/1/∞-system with (r, Q)-policy that will be investigated in Section 3. For comparison we also investigate the M/M/1/∞-system with (r, S)-policy in Section 4.

Theorem 2.2 *The continuous-time Markov process Z from Definition 2.1 is ergodic if and only if $\lambda < \mu$ and $\overline{p} < \infty$. If Z is ergodic then it has a unique limiting and stationary distribution of* *product form given by*

$$\pi(n,k) = K^{-1}\left(\frac{\lambda}{\mu}\right)^n q_k \ \text{with}\ n\in\mathbb{N}_0, k\geq 1, \tag{1}$$

$$\pi(n,0) = K^{-1}\left(\frac{\lambda}{\mu}\right)^n \frac{\lambda}{\nu} \ \text{with}\ n\in\mathbb{N}_0 \tag{2}$$

and with normalization constant

$$K = \frac{\mu}{\mu-\lambda}\left(\overline{p}+\frac{\lambda}{\nu}\right). \tag{3}$$

Proof. We first check that for finite K π is a stationary measure which satisfies the global balance equations of Z for all $n\in\mathbb{N}_0$:

$$\begin{aligned}&\pi(n,k)(\lambda+\mu(1-\delta_{0n}))\\ &= \pi(n-1,k)\lambda(1-\delta_{0n})+\pi(n+1,k+1)\mu+\pi(n,0)\nu p_k,\ k\geq 1,\end{aligned} \tag{4}$$

$$\begin{aligned}&\pi(n,0)\nu\\ &= \pi(n+1,1)\mu.\end{aligned} \tag{5}$$

This is done by insertion, using $q_1 = 1$ and $q_k = q_{k+1} + p_k$ for $k \geq 1$. Evaluating the normalization constant K yields the ergodicity criterion. □

We note further that the normalization constant K factorizes in the normalization constant for the marginal queue length and in that for the inventory process as follows $K = K_X \cdot K_Y$ with

$$K_X := \frac{\mu}{\mu-\lambda} \quad \text{and} \quad K_Y := \overline{p}+\frac{\lambda}{\nu}.$$

Remark 2.3 *The standard M/M/1/∞-FCFS system is obtained if $p_k = \delta_{k\infty}$ (Dirac measure concentrated in ∞). For $\nu=\infty$ the inventory is replenished instantaneously and the inventory position 0 is left immediately. Therefore, the stationary distribution is (1) and K simplifies to $K = \frac{\mu}{\mu-\lambda}\overline{p}$.*

Corollary 2.4 *(a) The marginal steady state queue length of $X = (X(t), t\geq 0)$ is equal to the steady state queue length distribution in the standard M/M/1/∞-FCFS system with the same parameters λ and μ.*
(b) The steady state on-hand inventory of $Y = (Y(t), t\geq 0)$ is

$$\mathsf{P}(Y=k) = \begin{cases} K_Y^{-1}\dfrac{\lambda}{\nu}, & \text{for } k=0\\[2ex] K_Y^{-1}q_k, & \text{for } k\geq 1.\end{cases}$$

Here we have denoted by Y a random variable distributed like the stationary inventory distribution.

Remark 2.5 *The result of Corollary 2.4 (a) is remarkable. λ is in effect not the arrival rate of customers to the system defined in Definition 2.1 and μ is not the actual service rate of customers. However, these rates must have been reduced to the same degree (in equilibrium) since if X has a steady state distribution of the form $\tilde{\pi}(n) = \rho^n(1-\rho)$, then $\rho = \frac{\lambda_{eff}}{\mu_{eff}}$ is the quotient of the effective input and service rates. Here $\rho = \frac{\lambda_{eff}}{\mu_{eff}} = \frac{\lambda}{\mu}$.*

Remark 2.6 *We have from Corollary 2.4 (b) that for $k \geq 1$*

$$\mathsf{P}(Y = k | Y > 0) = \frac{q_k}{\sum_{l=1}^{\infty} q_l} = \frac{\overline{p}^{-1} \sum_{h=k}^{\infty} p_h}{\sum_{l=1}^{\infty} \overline{p}^{-1} \sum_{h=l}^{\infty} p_h},$$

i.e. the stationary distribution of the inventory content, conditioned on having on-hand inventory greater than 0, is the conditional asymptotic and stationary residual interoccurence time in a discrete time renewal process with life time distribution $(p_h : h \geq 0)$. A comparison with the stationary residual service time for symmetric servers with lattice service time distributions reveals a deviation from the standard form (inventory position and residual service time are both supplementary variables), only the conditional stationary distribution of the supplementary variables coincide.

2.1 Measures of system performance

Having determined the stationary distribution, we can now compute several measures of operating characteristics for the system. We introduce the following measures of system performance for the stationary system: the average inventory position $\overline{I}$, the expected reorder rate λ_R, the expected lost sales per unit time $\overline{LS}$ and per cycle $\overline{LS}_c$, the safety stock s, α- and β-service levels and the average waiting time for a customer $\overline{W}$. Note that stationarity is always assumed in the classical inventory theory as well.
We say that the system goes through one cycle in the time between the placing of two successive orders or equivalently the receipt of two successive procurements with respect to the mean cycle time, i.e. the mean cycle time is λ_R^{-1}.

The stationary average on-hand inventory position is given by

$$\overline{I} = \sum_{k \geq 1} k \sum_{n \geq 0} \pi(n,k) = K_Y^{-1} \sum_{k \geq 1} k q_k. \tag{6}$$

Note that

$$\sum_{k \geq 1} k q_k = \sum_{k \geq 1} k \sum_{h=k}^{\infty} p_h = \sum_{k \geq 1} \left(\sum_{h=1}^{k} h \right) p_h = \frac{1}{2} \sum_{k \geq 1} k(k+1) p_k.$$

Hence, $\overline{I}$ depends on the first and second moment of the distribution F_p.

Using (66) of Theorem A.1 in Appendix A the expected number of replenishments per unit time is given by

$$\lambda_R = \frac{\lambda}{\overline{p} + \frac{\lambda}{\nu}}.$$

It is striking that λ_R is the same as the mean number of replenishment orders given by Hadley and Whitin [HW63] on page 180 in 4-37 and by Hax and Candea [HC84] in (4.1.38) for the case of no queueing (i.e. service time = 0). Saying it the other way round: The expected number of replenishments per unit time in the stationary system is independent of the service intensity as long as it is greater than the overall arrival rate λ. Clearly the same observation holds for the reciprocal value, the mean time between two replenishment orders $\dfrac{\overline{p}}{\lambda} + \dfrac{1}{\nu}$.

The average number of lost sales incurred per unit of time is given by

$$\overline{LS} = \lambda \mathsf{P}(Y = 0) = \frac{\lambda^2}{\overline{p}\nu + \lambda}.$$

The expected number of lost sales per cycle is given by

$$\overline{LS}_c = \frac{\overline{LS}}{\lambda_R} = \lambda/\nu.$$

Inventory can be devided into working stock and safety stock. Working stock is inventory acquired and held in advance of requirements so that the expected demand can be satisfied and ordering can be done on a lot size rather than on an as needed basis. Lot sizing is done in order to minimize ordering and holding costs. Safety stock is held in reserve to protect against the uncertainties of supply and demand (Tersine [Ter94] p. 205). The safety stock is usually defined as follows (see Silver/Peterson [SP85] Chapter 7).

Definition 2.7 *The safety stock is the Palm stationary mean value*

$$s = \mathsf{E}(\textit{ net inventory position just before the arrival of a replenishment order }).$$

We define the net inventory position as the inventory on hand

$$Y_{net} = Y.$$

The net inventory position is normally defined as the on-hand inventory less backorders. In our system customer demand is not backordered. Therefore, the above definition of net inventory is reasonable in the service facility with inventory and lost sales as defined in Definition 2.1 since the customer who is in service will not use one piece from inventory until he leaves the system. Another approach to evaluate the net inventory position Y_{net} will be investigated in Section 5. In the system under consideration the expected net inventory just before a replenishment order arrives is 0 but we shall need the general form of the safety stock s later in Sections 3 and 4.

Service levels are defined for a specified time period which can be the average replenishment cycle or one unit of time for example. Service levels based on the replenishment cycle simplify computations. To the customers, however, the length of an order cycle is of minor interest, they are interested in the quality of service in every unit of time. The definition of the α-service levels is a combination of the definitions given by [Sch81] and [SP85]. The definition of the β-service level is standard and can be found in [Sch81] and [SP85]. Note that the underlying probability measures may be casewise different.

Definition 2.8

$$\begin{aligned}
\alpha_1 &= \mathsf{P}\,(\textit{net inventory position at the end of a cycle } > 0),\\
\alpha_2 &= \mathsf{P}\,(\textit{net inventory position at the end of a unit of time } > 0),\\
\beta &= \frac{\mathsf{E}\,(\textit{demand satisfied per cycle})}{\mathsf{E}\,(\textit{total demand per cycle})} \stackrel{\star}{=} \frac{\mathsf{E}\,(\textit{demand satisfied per unit of time})}{\mathsf{E}\,(\textit{total demand per unit of time})}.
\end{aligned}$$

Here $\stackrel{\star}{=}$ is justified for all regenerative processes and therefore is here a conclusion of the ergodicity of Z. α-service levels are event-oriented performance measures. They only value the occurence but not the magnitude of a shortage. The definition of an α-service level depends on the chosen time interval, it represents the fraction of replenishment cycles or time units without a negative net inventory position. In the present case the α_2-service level is just the time stationary probability of a non-negative net inventory position. It is rarely considered in the literature. The β-service level is a quantity-oriented service measure describing the proportion of demands that are met from stock without accounting for the duration of a stockout. β-service levels are widely used in practice [Tem00].

According to Definition 2.7 of the net inventory position Y_{net} the α_1-service level is just 0 in the system defined in Definition 2.1 since the net inventory position is always 0 when an order arrives. (Note that if we use the definition from [Sch81] $\alpha_1 = \mathsf{P}\,(Y_{net}$ at the end of a cycle $\geq 0)$ then the α_1-service level will be equal to 1 for any M/M/1-system with inventory management and lost sales since with Definition 2.7 the net inventory position Y_{net} cannot become negative.) $\alpha_2 = \mathsf{P}(Y > 0)$. For the β-service level we compute

$$\beta = \frac{\mathsf{E}\,(\text{satisfied demand per unit of time})}{\mathsf{E}\,(\text{total demand per unit of time})} = \frac{\lambda - \overline{LS}}{\lambda} = \frac{\overline{p}}{\overline{p} + \overline{LS}_c} = 1 - \frac{\lambda}{\overline{p}\nu + \lambda}.$$

Hence

$$\beta = \mathsf{P}(Y > 0) = \alpha_2 \quad \text{if } Y_{net} = Y,$$

since the proportion of demand which is lost is just controlled by $Y = 0$. Normally, α- and β-service levels are not in a universally valid proportion to each other. Furthermore, since $\mathsf{P}(Y = 0) = \lambda_R/\nu$ we find

$$\lambda_R\,\overline{\alpha}_1 = \nu\,\overline{\alpha}_2,$$

where $\overline{\alpha}_i = 1 - \alpha_i$, $i = 1, 2$. In the classical inventory management literature the relation between α_1 and α_2 is $\lambda_R \overline{\alpha}_1 = \overline{\alpha}_2$.

α- and β-service levels do not provide any information on the length of the waiting time that a customer may experience. In contrast to these quantity related service measures, a second stream in the literature therefore uses a time criteria to measure inventory control performance. This approach becomes even more important in case of non-zero and stochastic service times, i.e. customer orders are not filled instantaneously from on-hand inventory as assumed in most inventory management systems. For example, Tempelmeier suggests in [Tem85] and [Tem00] to analyse an inventory model with a service constraint on the expected customer waiting time or on the probability that the customer waiting time is larger than a prespecified constant.

We assume that the waiting time of a customer ends at the moment he can enter the server since there are no more customers waiting in front of him. Thus, we exclude the replenishment lead time from the customer waiting time. From Corollary 2.4 (a) the mean number of customers in the system, $\overline{L}_0$, and the mean number of waiting customers, $\overline{L}$, are the same as in the standard M/M/1/∞-system. However, the mean waiting time of the customers is different from the expected waiting time in the M/M/1/∞-system since customers are not arriving with overall intensity λ because of the interruption of the arrival stream during the stochastic replenishment lead time. As can be found in Theorem A.1 in Appendix A the mean number of customers arriving per unit time is given by

$$\lambda_A = \overline{p}\lambda_R = \frac{\overline{p}\lambda\nu}{\overline{p}\nu + \lambda}.$$

Note that for the throughput of the system the following relation holds $\lambda_A = \lambda \mathsf{P}\,(Y > 0) = \lambda - \overline{LS}$.

Using Little's formula the mean sojourn time $\overline{W}_0$ and the mean waiting time $\overline{W}$ of the customers are given by

$$\overline{W}_0 = \frac{\overline{L}_0}{\lambda_A} = \frac{\overline{p}\nu + \lambda}{\overline{p}\nu(\mu - \lambda)} = \frac{1}{\mu - \lambda} + \frac{\lambda}{\overline{p}\nu(\mu - \lambda)}, \tag{7}$$

$$\overline{W} = \frac{\overline{L}}{\lambda_A} = \frac{(\overline{p}\nu + \lambda)\lambda}{\overline{p}\nu\mu(\mu - \lambda)} = \frac{\lambda}{\mu(\mu - \lambda)} + \frac{\lambda^2}{\overline{p}\nu\mu(\mu - \lambda)}. \tag{8}$$

$\overline{W}_0$ and $\overline{W}$ are naturally larger than the mean sojourn time and mean waiting time of the standard M/M/1/∞-system, respectively, which are just the first summands of the above expressions. The second term is $\lambda/(\overline{p}\nu)$ times the first summand. These characteristics of $\overline{W}_0$ and $\overline{W}$ just result from the mean interarrival time of customers who are admitted to the system, $\lambda_A^{-1} = \lambda^{-1} + (\overline{p}\nu)^{-1}$, which is larger than in the standard M/M/1/∞-system.

Remark 2.9 *Coming back to the special cases discussed in Remark 2.3, note that if $\overline{p} \to \infty$ the system behaves like the standard M/M/1/∞-system with the same $\lambda_A = \lambda$, $\overline{W}_0$ and $\overline{W}$. The inventory size and the cycle length are infinite, no demand is lost and $\beta = 1$. If $\nu \to \infty$ then λ_A, $\overline{W}_0$ and $\overline{W}$ are the same as in the standard M/M/1/∞-system, no demand is lost and $\beta = 1$ as well. However, in this case the cycle length is finite: $\lambda_R = \lim_{\nu\to\infty} \lambda(\overline{p} + \frac{\lambda}{\nu})^{-1} = \lambda/\overline{p}$.*

Remark 2.10 *Note that only $\overline{W}_0$ and $\overline{W}$ depend on the service rate μ. Some performance measures are not dependent on λ and ν individually but only on their proportion λ/ν, e.g. $\overline{I}$, $\overline{LS}_c$ and β. Concerning the influence of F_p we observe that several performance measures only depend on the first moment of F_p like $\lambda_R, \overline{LS}, \beta$ and $\overline{W}$ or are completely independent of F_p like $\overline{LS}_c$ and α_1. $\overline{I}$ depends on the first and second moment of F_p.*
Hence, for two systems which have the same parameters λ, ν and μ but different order size distributions F_p and $F_{\tilde{p}}$ with the same mean $\overline{p}$ only $\overline{I}$ will be different.

2.2 Some examples for the order size distribution

In this section we will investigate three examples of the replenishment order size distribution. We consider the fixed order size system with inventory size Q, which yields an $(0, Q)$-policy in the traditional terminology, the system with geometrically distributed order size and the system with uniformly distributed order sizes on $\{1, ..., Q\}$.

Deterministic order size. Let us assume that the order size is fixed and equal to $Q \in \mathbb{N}$. Hence, we have $p_k = \delta_{kQ}$ for all $k \in \mathbb{N}$. Then $q_k = 1$ for all $k \in 1, ..., Q$ and $q_k = 0$ otherwise. The mean order size is given by $\overline{p} = Q$ and the normalization constant K takes on the form

$$K = \frac{\mu}{\mu - \lambda}\left(Q + \frac{\lambda}{\nu}\right).$$

Note that as in Berman and Sapna's [BS00] paper the steady state probabilities for the on-hand inventory being $k \geq 1$ are equally distributed according to

$$\mathsf{P}(Y = k) = \left(Q + \frac{\lambda}{\nu}\right)^{-1}, \quad 1 \leq k \leq Q.$$

In contrast to their results the probability that a replenishment order is outstanding here has a different probability. It is given by

$$\mathsf{P}(Y = 0) = \frac{\lambda}{\nu}\left(Q + \frac{\lambda}{\nu}\right)^{-1}.$$

Table 1: Performance measures

F_p	Deterministic	Geometric	Uniform
$\overline{I}$	$\frac{Q}{Q+\frac{\lambda}{\nu}} \frac{Q+1}{2}$	$\frac{\nu}{(1-q)\nu-(1-q)^2\lambda}$	$\frac{(Q+2)[2\lambda+(Q+1)\nu]}{6\lambda}$
λ_R	$\frac{\lambda}{Q+\frac{\lambda}{\nu}}$	$\frac{(1-q)\lambda\nu}{\nu+(1-q)\lambda}$	$\frac{2\lambda\nu}{(Q+1)\nu+2\lambda}$
$\overline{LS}$	$\frac{\lambda^2}{Q\nu+\lambda}$	$\frac{(1-q)\lambda^2}{\nu+(1-q)\lambda}$	$\frac{2\lambda^2}{(Q+1)\nu+2\lambda}$
β	$\frac{Q}{Q+\frac{\lambda}{\nu}}$	$\frac{\nu}{\nu+(1-q)\lambda}$	$\frac{(Q+1)\nu}{(Q+1)\nu+2\lambda}$
$\overline{W}$	$\frac{(Q\nu+\lambda)\lambda}{Q\nu\mu(\mu-\lambda)}$	$\frac{(\nu+(1-q)\lambda)\lambda}{\nu\mu(\mu-\lambda)}$	$\frac{[(Q+1)\nu+2\lambda]\lambda}{(Q+1)\nu\mu(\mu-\lambda)}$

Geometrically distributed order size. Let the size of a replenishment order be geometrically distributed with parameter q ($\text{geo}^{(+)}(q)$). Thus, $p_k = q^{k-1}(1-q)$ and $q_k = \sum_{h=k}^{\infty} p_k = q^{k-1}$ for all $k \geq 1$. Using $\overline{p} = \sum_{h=1}^{\infty} hp_h = (1-q)^{-1}$ we obtain

$$K \;=\; \frac{\mu}{\mu-\lambda}\left(\frac{1}{1-q}+\frac{\lambda}{\nu}\right).$$

Uniformly distributed order size. Let the size of a replenishment order be equally distributed on $\{1,...,Q\}$ ($\text{Unif}(\{1,...,Q\})$), hence $p_k = 1/Q$, for all $k \in \{1,...,Q\}$ and $p_k = 0$ otherwise. Here $\overline{p} = \frac{Q+1}{2}$ and $q_k = \sum_{h=k}^{Q} p_h = \frac{Q+1-k}{Q}$. The normalization constant K is given by

$$K = \frac{\mu}{\mu-\lambda}\left(\frac{Q+1}{2}+\frac{\lambda}{\nu}\right).$$

The performance measures of these three examples are summarized in Table 1.

2.3 Ordering result

Consider two single server systems as in Definition 2.1 which differ only in having different probability distribution functions F_p and $F_{\tilde{p}}$ for the size of replenishment orders. The mean order sizes in the two systems are assumed to be the same. Denote by $X, \tilde{X}$ and $Y, \tilde{Y}$ the random variables distributed like the stationary queue length and inventory distribution of the two systems, respectively. Then, the following theorem holds

Theorem 2.11 *If $F_p <_{cx} F_{\tilde{p}}$ then $(X,Y) <_{st} (\tilde{X},\tilde{Y})$.*

Proof. From Theorem A in [Sze95] section 1.3, p. 10 $F_p <_{cx} F_{\tilde{p}}$ is equivalent to $\sum_{h=k}^{\infty} \overline{F}_p(k) \leq \sum_{h=k}^{\infty} \overline{F}_{\tilde{p}}(k)$ for all $k \in \mathbb{N}$. This is $\sum_{h=k}^{\infty} q_k \leq \sum_{h=k}^{\infty} \tilde{q}_k$ for all $k \in \mathbb{N}$. Hence, $Y <_{st} \tilde{Y}$ since $\mathsf{P}(Y > k) \leq \mathsf{P}(\tilde{Y} > k)$ holds for all $k \in \mathbb{N}$ and $k = 0$. Moreover, because steady state probabilities are in product form (1) and λ, μ and ν are the same in both systems, the result follows for $k \in \mathbb{N}_0$ ($\mathsf{P}(X = n, Y > k) \leq \mathsf{P}(\tilde{X} = n, \tilde{Y} > k)$ for all $n \in \mathbb{N}_0$ and $k \in \mathbb{N}_0$). Thus, $(X,Y) <_{st} (\tilde{X},\tilde{Y})$. $\square$

Remark 2.12 *The convex ordering of two distribution functions F_p and $F_{\tilde{p}}$, with finite and equal first moments can be ascertained using the Karlin-Novikoff cut-criterion (see for example [Sze95] section 1.3, Theorem E, p. 17). Suppose that for F_p and $F_{\tilde{p}}$ we have*

$$\begin{aligned} F_p(k) &\leq F_{\tilde{p}}(k), \text{ for } k \leq \xi, \\ F_p(k) &\geq F_{\tilde{p}}(k), \text{ for } k > \xi, \end{aligned}$$

for some $\xi \in \mathbb{R}_+$ then

$$F_p <_{cx} F_{\tilde{p}}.$$

Example 2.13 *Let $p \sim \delta_Q$, then $F_p(k) = 0$ for all $k < Q$ and $F_p(k) = 1$ for all $k \geq Q$ and $\overline{p} = Q$. Hence, the Karlin-Novikoff cut criterion can be applied for comparison with all other distribution functions $F_{\tilde{p}}$ with mean Q. Two examples which have already been studied above are*

1. $p \sim \delta_{1/(1-q)} <_{cx} \tilde{p} \sim geo^+(q)$,
2. $p \sim \delta_{(Q+1)/2} <_{cx} \tilde{p} \sim$ *Unif*$(\{1, ..., Q\})$.

Consequently, if the mean replenishment size is prescribed, the stationary mean inventory position $\overline{I}$ is minimal for the system with fixed order size. This can be seen as follows: Consider two systems with different probability functions p and $\tilde{p}$, with $p \sim \delta_Q$ and $\mathsf{E}\tilde{p} = Q$. Then $p <_{cx} \tilde{p}$. If we denote by Y and $\tilde{Y}$ the random variables distributed like the stationary inventory distribution of the two systems it follows from Theorem 2.11 that $Y <_{st} \tilde{Y}$. Hence,

$$\overline{I} = \sum_{k \in \mathbb{N}} \mathsf{P}(Y \geq k) \leq \sum_{k \in \mathbb{N}} \mathsf{P}(\tilde{Y} \geq k) = \overline{\tilde{I}}.$$

We have found out that all performance measures except the mean inventory position $\overline{I}$ are the same for two systems that differ only in the order size distribution function F_p but have the same mean $\overline{p}$. $\overline{I}$ is minimal for the fixed order size system from Section 2.2. This is why all other systems can be excluded from further consideration when it comes to an optimization of the costs associated to M/M/1/∞-systems with inventory management and reorder point 0.
For the same reason we will now investigate early replenishment systems with reorder point $r \geq 0$ and fixed order size Q. This results in the well-known (r, Q)-system from inventory management theory.

3 M/M/1/∞-system with (r, Q)-policy

In this section we consider an M/M/1/∞-system with inventory management policy that corresponds to the lost sales version of a continuous review, order-point, order-quantity model, typically called (r, Q) model, where a fixed order quantity Q is ordered each time the on-hand inventory reaches the reorder point r (see [Ter94] p. 206 and [SP85] p. 256). The (r, Q)-policy is applicable if units are demanded one at a time and if transaction reporting is used. We assume throughout the paper that $r < Q$. This avoids degenerated cycles in which no demand occurs.

Definition 3.1 *An M/M/1/∞-system with (r, Q)-policy is a single server with attached inventory and infinite waiting room under FCFS regime as in Definition 2.1. If the on-hand inventory reaches a prespecified value $r \geq 0$, a replenishment order is instantaneously triggered. The size of the replenishment order is fixed to Q units, $Q > r$. The replenishment lead time is exponentially distributed with parameter ν, $\nu > 0$. During the time the inventory is depleted, arriving customers are lost.*

Let $X(t)$ denote the number of customers present at the server at time $t \geq 0$, either waiting or in service and let $Y(t)$ denote the on-hand inventory at time $t \geq 0$. Then $Z = ((X(t), Y(t)), t \geq 0)$ is a continuous-time Markov process for the M/M/1/∞-system with (r,Q)-policy. The state space of Z is $E_Z = \{(n,k) : n \in \mathbb{N}_0, 0 \leq k \leq Q + r\}$.

Remark 3.2 *Note that as in Section 2 an infinite order size $Q = \infty$ results in the standard M/M/1/∞-system. If $\nu = \infty$ an early replenishment of the inventory with $r > 0$ does not make sense with respect to economical aspects, since r units are never touched by the customers and remain in stock for ever.*

Theorem 3.3 *The continuous-time Markov process Z from Definition 3.1 is ergodic if and only if $\lambda < \mu$ and $Q, \nu < \infty$. If Z is ergodic then it has a unique limiting and stationary distribution of product form given by*

$$\pi(n,k) = K^{-1} C(k) \left(\frac{\lambda}{\mu}\right)^n, \qquad \text{for } n \in \mathbb{N}_0, 1 \leq k \leq Q + r, \tag{9}$$

$$\pi(n,0) = K^{-1} \frac{\lambda}{\nu} \left(\frac{\lambda}{\mu}\right)^n, \qquad \text{for } n \in \mathbb{N}_0, \tag{10}$$

with

$$\begin{aligned} C(k) &= \left(\frac{\lambda+\nu}{\lambda}\right)^{k-1}, \quad k = 1, ..., r, \\ C(k) &= \left(\frac{\lambda+\nu}{\lambda}\right)^{r}, \quad k = r+1, ..., Q, \\ C(Q+k) &= \left(\frac{\lambda+\nu}{\lambda}\right)^{r} - \left(\frac{\lambda+\nu}{\lambda}\right)^{k-1}, \quad k = 1, ..., r. \end{aligned}$$

The normalization constant K is given by

$$K = \frac{\mu}{\mu - \lambda} \left(Q \left(\frac{\lambda+\nu}{\lambda}\right)^r + \frac{\lambda}{\nu} \right). \tag{11}$$

Proof. The process Z from Definition 3.1 possesses the following global balance equations:

$$\begin{aligned} &\pi(n,k)(\lambda + \mu(1 - \delta_{0n}) + \nu 1_{\{k \leq r\}}) \\ &= \pi(n-1,k)\lambda(1-\delta_{0n}) + \pi(n+1,k+1)\mu(1-\delta_{Q+r,k}) \\ &\quad + \pi(n,k-Q)\nu 1_{\{k \geq Q\}}, \quad 1 \leq k \leq Q + r, \end{aligned} \tag{12}$$

$$\begin{aligned} &\pi(n,0)\nu \\ &= \pi(n+1,1)\mu \end{aligned} \tag{13}$$

for all $n \in \mathbb{N}_0$. We will show that the distribution from (9) and (10) satisfies equations (12) and (13). Equation (13) is satisfied since $C(1) = 1$. In (12) the term $\pi(n,k)\mu(1-\delta_{0n})$ at the left cancels against the term $\pi(n-1,k)\lambda(1-\delta_{0n})$ at the right side for all $n \in \mathbb{N}_0$ and $1 \leq k \leq Q+r$. It remains to prove that the term on the left side

$$\pi(n,k)(\lambda + \nu 1_{\{k \leq r\}}) = K^{-1} \left(\frac{\lambda}{\mu}\right)^n C(k)(\lambda + \nu 1_{\{k \leq r\}})$$

equals the remaining terms at the right side

$$\begin{aligned}
&\pi(n+1,k+1)\mu(1-\delta_{Q+r,k})+\pi(n,k-Q)\nu 1_{\{k\geq Q\}}\\
&= K^{-1}\left(\frac{\lambda}{\mu}\right)^n \lambda C(k+1)(1-\delta_{Q+r,k})+K^{-1}\left(\frac{\lambda}{\mu}\right)^n \lambda 1_{\{k=Q\}}\\
&\quad +K^{-1}\left(\frac{\lambda}{\mu}\right)^n \nu C(k-Q)1_{\{k>Q\}}
\end{aligned}$$

for all $n\in\mathbb{N}_0$ and $1\leq k\leq Q+r$. Hence, we have to show that for all $1\leq k\leq Q+r$

$$C(k+1)(1-\delta_{Q+r,k}) = C(k)\left(1+\frac{\nu}{\lambda}1_{\{k\leq r\}}\right)-1_{\{k=Q\}}-\frac{\nu}{\lambda}C(k-Q)1_{\{k>Q\}} \tag{14}$$

holds. The equations which have to be satisfied are given by

$$\begin{aligned}
C(k+1) &= C(k)\left(1+\frac{\nu}{\lambda}\right), \quad k=1,...,r,\\
C(k+1) &= C(k), \qquad k=r+1,...,Q-1,\\
C(Q+1) &= C(Q)-1,\\
C(k+1) &= C(k)-\frac{\nu}{\lambda}C(k-Q), \quad k=Q+1,...,Q+r-1,\\
0 &= C(Q+r)-\frac{\nu}{\lambda}C(r).
\end{aligned}$$

The first Q equations are satisfied by the $C(k)$ given in Theorem 3.3. For $k\geq 1$ we show by induction that if $C(Q+1)=C(Q)-C(1)$ and $C(Q+k+1)=C(Q+k)-\frac{\nu}{\lambda}C(k)$ it follows that $C(Q+k)=C(Q)-C(k)$. First note that from the first equation $C(Q+1)=C(Q)-C(1)$ the induction start is done. Assume that for some $1\leq k<r$ the equality $C(Q+k)=C(Q)-C(k)$ holds. Then, we obtain for $k+1$: $C(Q+k+1)=C(Q+k)-\frac{\nu}{\lambda}C(k)=C(Q)-C(k)-\frac{\nu}{\lambda}C(k)=C(Q)-C(k+1)$. The last equation is satisfied since $C(Q+r)=C(Q)-C(r)=\left(\frac{\lambda+\nu}{\lambda}-1\right)C(r)=\frac{\nu}{\lambda}C(r)$.
The normalization constant K is given by

$$\begin{aligned}
K &= \sum_{n\geq 0}\sum_{k=1}^{Q+r}\left(\frac{\lambda}{\mu}\right)^n C(k)+\sum_{n\geq 0}\left(\frac{\lambda}{\mu}\right)^n\frac{\lambda}{\nu}=\frac{\mu}{\mu-\lambda}\left(\sum_{k=1}^{Q+r}C(k)+\frac{\lambda}{\nu}\right)\\
&= \frac{\mu}{\mu-\lambda}\left(Q\left(\frac{\lambda+\nu}{\lambda}\right)^r+\frac{\lambda}{\nu}\right). \qquad \square
\end{aligned}$$

As in Section 2 the normalization constant K factorizes into

$$K_X := \frac{\mu}{\mu-\lambda} \quad \text{and} \quad K_Y := \left(Q\left(\frac{\lambda+\nu}{\lambda}\right)^r+\frac{\lambda}{\nu}\right).$$

The steady state queue length of $X=(X(t),t\geq 0)$ is equally distributed as the steady state queue length in the standard M/M/1/∞-FCFS system with the same parameters λ and μ. Especially note that the inventory positions in steady state are not equally distributed on $(s,Q+s]$ (where s is the safety stock) as assumed by Zheng [Zhe92] for example.

Remark 3.4 *Buchanan and Love [BL85] investigate an (r,Q) inventory system with lost sales and Erlang-distributed lead times, where customer demands are satisfied without any loss of time. Their result for the special case of exponentially distributed lead times coincides with the marginal distribution of the inventory process of our system.*

Theorem 3.5 (Measures of system performance) *The measures of system performance as defined in Section 2.1 of the M/M/1/∞-system with (r,Q)-policy according to Definition 3.1 are given by*

$$\overline{I} = \frac{Q}{Q+\frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}\left(\frac{Q+1}{2}+s\right), \tag{15}$$

$$\lambda_R = \frac{\lambda}{Q+\frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}, \tag{16}$$

$$\lambda_A = \frac{\lambda Q}{Q+\frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}, \tag{17}$$

$$\overline{LS} = \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r \frac{\lambda}{Q+\frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}, \tag{18}$$

$$\overline{LS}_c = \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r, \tag{19}$$

$$s = r-\frac{\lambda}{\nu}\left(1-\left(\frac{\lambda}{\lambda+\nu}\right)^r\right), \tag{20}$$

$$\alpha_1 = 1-\left(\frac{\lambda}{\lambda+\nu}\right)^r, \tag{21}$$

$$\alpha_2 = \beta = \frac{Q}{Q+\frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}, \tag{22}$$

$$\overline{W}_0 = \frac{1}{\mu-\lambda}+\frac{\lambda}{Q\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r\frac{1}{\mu-\lambda}, \tag{23}$$

$$\overline{W} = \frac{\lambda}{\mu(\mu-\lambda)}+\frac{\lambda}{Q\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r\frac{\lambda}{\mu(\mu-\lambda)}. \tag{24}$$

Proof. The proof of Theorem 3.5 is presented in Appendix B on page 28. ☐

Some remarks may be in order here. Obviously, $\overline{I} < \frac{Q+1}{2}+s$. The expected number of replenishments per unit time λ_R has the same form as in Section 2.1 (mean demand per unit time/(mean satisfied demand per cycle + mean lost sales per cycle)) and $\lambda_A = \lambda_R\cdot$ (mean satisfied demand per cycle).

The safety stock s is equal to the difference of the reorder point r and the expected lead time demand which can be satisfied. $s=0$ for $r=0$ and s is strictly increasing in the reorder point r since for all $r \geq 1$:

$$s(r)-s(r-1) = 1-\frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^{r-1}\left(1-\frac{\lambda}{\lambda+\nu}\right) = 1-\left(\frac{\lambda}{\lambda+\nu}\right)^r > 0. \tag{25}$$

α_1 can be interpreted as the proportion of cycles without a stockout or as the proportion of lead time demand which can be satisfied from stock. This explains the relation $s = r-\frac{\lambda}{\nu}\alpha_1$ once again. α_1 is strictly increasing in r.

As before $\beta = \mathsf{P}(Y>0) = \alpha_2 = \lambda_A/\lambda$. $\overline{W}$ and $\overline{W}_0$ are again larger than in the standard M/M/1/∞-system, since $\overline{W} = \overline{W}_0\frac{\lambda}{\mu} = \frac{\lambda}{\mu(\mu-\lambda)}\left(1+\frac{\overline{LS}_c}{Q}\right)$.

Remark 3.6 *Remark 2.9 carries over to the cases $Q \to \infty$ and $\nu \to \infty$ in the M/M/1/∞-system with (r, Q)-policy as defined in Definition 3.1. Furthermore, if $\nu \to \infty$ the mean inventory on hand is $(Q+1)/2 + r$ and r units remain in stock for ever.*

Remark 3.7 *As in Section 2 the only performance measures that depend on the service rate μ are $\overline{W}_0$ and $\overline{W}$. The values for $\overline{I}, \overline{LS}_c, s$ and β are not dependent on λ and ν individually but only on their proportion λ/ν. Concerning the influence of Q and r we observe that all performance measures depend on the reorder rate r, but only $\lambda_R, \lambda_A, \overline{LS}, \beta, \overline{W}_0$ and $\overline{W}$ are also dependent on Q.*

4 M/M/1/∞-system with (r, S)-policy

The M/M/1/∞-system with inventory management studied in this section corresponds to the lost sales version of a continuous review, order-point, order-up-to-level (r, S)-system. Each time the on-hand inventory reaches the reorder point r a variable replenishment quantity is used such that upon replenishment, the on-hand inventory is restocked to level S (see [BS01] and [SSL92] for an equivalent definition of the (r, S)-policy in the backorder case). In the inventory management literature another definition of the (r, S)-system has appeared as well (see [SP85] p. 256, [HW63]), where the order size is determined at the moment the order is placed such that the actual inventory level $x < r$ plus the variable order size is equal to S. In our case, where demands are unit-sized, no overshoot of the order point r is possible and the (r, Q)- and (r, S)-systems are identical when the second definition is used.

Definition 4.1 *An M/M/1/∞-system with (r, S)-policy is a single server with attached inventory and infinite waiting room under FCFS regime as in Definition 2.1. If the inventory reaches a prespecified value $r > 0$, a replenishment order is instantaneously triggered. With each replenishment the inventory level is restocked to S units with $r < S$. The replenishment lead time is exponentially distributed with parameter ν, $\nu > 0$. During the time the inventory is depleted, no customers are admitted to join the queue and are lost.*
Let $X(t)$ denote the number of customers present at the server at time $t \geq 0$, either waiting or in service and let $Y(t)$ denote the on-hand inventory at time $t \geq 0$. Then $Z = ((X(t), Y(t)), t \geq 0)$ is a continuous-time Markov process for the M/M/1/∞-system with (r, S)-policy. The state space of Z is given by $E_Z = \{(n, k) : n \in \mathbb{N}_0, 0 \leq k \leq S\}$.

Remark 4.2 *Note that as in Sections 2 and 3 an infinite order-up-to-level $S = \infty$ results in the standard M/M/1/∞-system and an early replenishment of the inventory with $r > 0$ does not make sense if $\nu = \infty$.*

Theorem 4.3 *The continuous-time Markov process Z from Definition 4.1 is ergodic if and only if $\lambda < \mu$ and $S, \nu < \infty$. If Z is ergodic then it has a unique limiting and stationary distribution of product form given by*

$$\pi(n, k) = K^{-1} \left(\frac{\lambda}{\mu}\right)^n C(k), \qquad \text{for } n \in \mathbb{N}_0, 1 \leq k \leq S, \tag{26}$$

$$\pi(n, 0) = K^{-1} \left(\frac{\lambda}{\mu}\right)^n \frac{\lambda}{\nu}, \qquad \text{for } n \in \mathbb{N}_0, \tag{27}$$

with

$$\begin{aligned} C(k) &= \left(\frac{\lambda+\nu}{\lambda}\right)^{k-1}, \ k = 1, ..., r, \\ C(k) &= \left(\frac{\lambda+\nu}{\lambda}\right)^{r}, \quad k = r+1, ..., S \end{aligned}$$

and normalization constant

$$K = \frac{\mu}{\mu-\lambda}\left(S - r + \frac{\lambda}{\nu}\right)\left(\frac{\lambda+\nu}{\lambda}\right)^r. \tag{28}$$

Proof. The process Z from Definition 4.1 possesses the following global balance equations:

$$\begin{aligned}
&\pi(n,k)(\lambda + \mu(1-\delta_{0n}) + \nu 1_{\{k\le r\}}) \\
&\quad = \pi(n-1,k)\lambda(1-\delta_{0n}) + \pi(n+1,k+1)\mu(1-\delta_{Sk}) \\
&\qquad + \sum_{h=0}^{r}\pi(n,h)\nu 1_{\{k=S\}}, \quad 1 \le k \le S,
\end{aligned} \tag{29}$$

$$\begin{aligned}
&\pi(n,0)\nu \\
&\quad = \pi(n+1,1)\mu
\end{aligned} \tag{30}$$

for all $n \in \mathbb{N}_0$. We will show that the distribution from (26) and (27) satisfies equations (29) and (30). Equation (30) is satisfied since $C(1) = 1$. In (29) the term $\pi(n,k)\mu(1-\delta_{0n})$ at the left cancels against the term $\pi(n-1,k)\lambda(1-\delta_{0n})$ at the right side for all $n \in \mathbb{N}_0$ and $1 \le k \le S$. Moreover, we will prove that the term on the left side

$$\pi(n,k)(\lambda + \nu 1_{\{k\le r\}}) = K^{-1}\left(\frac{\lambda}{\mu}\right)^n C(k)(\lambda + \nu 1_{\{k\le r\}})$$

corresponds to the following expression at the right side

$$\begin{aligned}
&\pi(n+1,k+1)\mu(1-\delta_{Sk}) + \sum_{h=0}^{r}\pi(n,h)\nu 1_{\{k=S\}} \\
&= K^{-1}\left(\frac{\lambda}{\mu}\right)^n \lambda C(k+1)(1-\delta_{Sk}) + K^{-1}\left(\frac{\lambda}{\mu}\right)^n \lambda 1_{\{k=S\}} \\
&\quad + K^{-1}\left(\frac{\lambda}{\mu}\right)^n \nu \sum_{h=1}^{r} C(h) 1_{\{k=S\}}
\end{aligned}$$

for all $n \in \mathbb{N}_0$ and $1 \le k \le S$. We therefore have to show that for all $1 \le k \le S$

$$C(k+1)(1-\delta_{S,k}) = C(k)\left(1 + \frac{\nu}{\lambda}1_{\{k\le r\}}\right) - 1_{\{k=S\}} - \frac{\nu}{\lambda}\sum_{h=1}^{r} C(h)1_{\{k=S\}} \tag{31}$$

holds. Hence, the equations which have to be satisfied by $C(k), 1 \le k \le S$ are given by

$$\begin{aligned}
C(k+1) &= C(k)\left(1 + \frac{\nu}{\lambda}\right), && k = 1, ..., r, \\
C(k+1) &= C(k), && k = r+1, ..., S-1, \\
0 &= C(S) - 1 - \frac{\nu}{\lambda}\sum_{h=1}^{r} C(h).
\end{aligned}$$

Obviously, the first $S-1$ equations are satisfied by the $C(k)$ as defined in Theorem 4.3. For the last equation we have $1 + \frac{\nu}{\lambda}\sum_{h=1}^{r} C(h) = 1 + \frac{\nu}{\lambda}\sum_{h=0}^{r-1}\left(\frac{\lambda+\nu}{\lambda}\right)^h = 1 - \left(1 - \left(\frac{\lambda+\nu}{\lambda}\right)^r\right) = C(S)$. Putting this together equation (29) is satisfied.

The normalization constant K is given by

$$
\begin{aligned}
K &= \sum_{n\geq 0}\sum_{k=1}^{S}\left(\frac{\lambda}{\mu}\right)^n C(k) + \sum_{n\geq 0}\left(\frac{\lambda}{\mu}\right)^n \frac{\lambda}{\nu} = \frac{\mu}{\mu-\lambda}\left(\sum_{k=1}^{S} C(k) + \frac{\lambda}{\nu}\right) \\
&= \frac{\mu}{\mu-\lambda}\left(\frac{1}{\lambda^{r-1}}\frac{\lambda^r - (\lambda+\nu)^r}{-\nu} + (S-r)\left(\frac{\lambda+\nu}{\lambda}\right)^r + \frac{\lambda}{\nu}\right) \\
&= \frac{\mu}{\mu-\lambda}\left(S - r + \frac{\lambda}{\nu}\right)\left(\frac{\lambda+\nu}{\lambda}\right)^r .
\end{aligned}
$$

□

The normalization constant K can be splitted up into two parts

$$
K_X := \frac{\mu}{\mu-\lambda} \quad \text{and} \quad K_Y := \left(S - r + \frac{\lambda}{\nu}\right)\left(\frac{\lambda+\nu}{\lambda}\right)^r .
$$

As in Sections 2 and 3 the steady state queue length distribution of $X = (X(t), t \geq 0)$ is equal to the steady state queue length distribution in the standard M/M/1/∞-FCFS system with the same parameters λ and μ.

Theorem 4.4 (Measures of system performance) *The measures of system performance as defined in Section 2.1 of the M/M/1/∞-system with (r,S)-policy according to Definition 4.1 are given by*

$$
\overline{I} = \frac{1}{S - r + \frac{\lambda}{\nu}}\left\{\frac{\lambda}{\nu}s + \frac{(S+1)S - (r+1)r}{2}\right\}, \tag{32}
$$

$$
\lambda_R = \frac{\lambda}{S - r + \frac{\lambda}{\nu}}, \tag{33}
$$

$$
\lambda_A = \lambda - \frac{\lambda^2}{(S-r)\nu + \lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r, \tag{34}
$$

$$
\overline{LS} = \frac{\lambda^2}{(S-r)\nu + \lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r, \tag{35}
$$

$$
\overline{LS}_c = \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r, \tag{36}
$$

$$
s = r - \frac{\lambda}{\nu}\left(1 - \left(\frac{\lambda}{\lambda+\nu}\right)^r\right), \tag{37}
$$

$$
\alpha_1 = 1 - \left(\frac{\lambda}{\lambda+\nu}\right)^r, \tag{38}
$$

$$
\alpha_2 = \beta = 1 - \frac{\lambda}{(S-r)\nu + \lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r, \tag{39}
$$

$$
\overline{W}_0 = \frac{1}{\mu-\lambda} + \frac{\lambda}{(S-s)\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r \frac{1}{\mu-\lambda}, \tag{40}
$$

$$
\overline{W} = \frac{\lambda}{\mu(\mu-\lambda)} + \frac{\lambda}{(S-s)\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r \frac{\lambda}{\mu(\mu-\lambda)}. \tag{41}
$$

Proof. The proof of Theorem 4.4 is presented in Appendix B on page 29. □

Remark 4.5 *The dependences of the performance measures on the several parameters are the same as in Section 3.*

Theorem 4.6 (Comparison of (r, Q)- and (r, S)-policy with the same r) *Let $S = Q + s$, where s is the safety stock from (20). Then, the stationary mean inventory position in the M/M/1/∞-system with (r, S)-policy is not larger than the stationary mean inventory position in the M/M/1/∞-system with (r, Q)-policy. All other performance measures are the same.*

Proof. The proof is given in Appendix B on page 30. □

5 Net inventory revisited

In Definition 2.7 of Section 2.1 we defined the net inventory position as the inventory on hand $Y_{net} = Y$. In the context of service facilities with inventory and lost sales a second common-sense definition of the net inventory position is

$$Y_{net} = \max\{0, Y - X\}. \tag{42}$$

This definition is reasonable since one can argue that for each customer in queue who has been admitted to the system one piece from the inventory is already reserved.

In the following theorems the performance measures dependent on Y_{net}, the safety stock s and the α_1-service level, will be given for all the inventory models considered above using the second definition of Y_{net} from (42).

Theorem 5.1 *In the M/M/1/∞-system with inventory management, reorder point 0 and arbitrary size of replenishment orders according to Definition 2.1, it follows*

$$s = 0 \quad \textit{and} \quad \alpha_1 = 0,$$

if the net inventory position is defined by (42).

Proof. See Appendix B, page 33. □

Theorem 5.2 *In the M/M/1/∞-system with (r, Q)- or (r, S)-policy according to Definition 3.1, resp. 4.1, it follows*

$$s = r - \frac{\lambda}{\nu}\left(1 - \left(\frac{\lambda}{\lambda+\nu}\right)^r\right) - \frac{\lambda}{\mu-\lambda}\alpha_1$$

and

$$\alpha_1 = 1 - \frac{1}{\mu-(\lambda+\nu)}\left((\mu-\lambda)\left(\frac{\lambda}{\lambda+\nu}\right)^r - \nu\left(\frac{\lambda}{\mu}\right)^r\right),$$

if the net inventory position is defined by (42).

Proof. The proof is given in Appendix B on page 33. □

We observe that α_1 and s now depend on the service rate μ. α_1 can be interpreted as the proportion of cycles where at the end of the cycle the number of customers is smaller than the size of the on-hand inventory. $\lambda/(\mu - \lambda)$ is the mean number of customers in the system. Hence, s is equal to the reorder point r minus the lead time demand which can be satisfied from stock minus α_1 times the mean number of customers in the system.

Corollary 5.3 *In the M/M/1/∞-system with (r, Q)- or (r, S)-policy according to Definition 3.1, resp. 4.1, α_1 has the following properties if the net inventory position is defined as in (42):*

1. $0 \leq \alpha_1 \leq 1 - \left(\frac{\lambda}{\mu}\right)^r$ *for some fixed* $r \in \mathbb{N}_0$,

2. $\alpha_1(\cdot)$ *is strictly increasing in* $r \in \mathbb{N}_0$.

Proof. See Appendix B, page 34. $\square$

Corollary 5.4 *In the M/M/1/∞-system with (r, Q)- or (r, S)-policy according to Definition 3.1, resp. 4.1, with net inventory position defined as in (42) the following equality holds*

$$s \leq r - \frac{\lambda}{\nu}\left(1 - \left(\frac{\lambda}{\lambda+\nu}\right)^r\right).$$

Proof. As shown in Corollary 5.3 it holds for all $r \in \mathbb{N}_0$ that $\alpha_1 \geq 0$ and α_1 is increasing in r. $\mu > \lambda$ follows from the ergodicity criterion, thus, the second summand of s is non-positive and decreasing in r. This completes the proof. $\square$

Remark 5.5 *If the net inventory position is defined by (42) and we use the definition of α_1 from [Sch81] denoted by $\alpha_{S,1}$, then for the M/M/1/∞-system with inventory management, reorder point 0 and arbitrary size of replenishment orders according to Definition 2.1*

$$\alpha_{S,1} := \mathsf{P}(Y_{net} \textit{ at the end of a cycle } \geq 0) = \frac{\mu - \lambda}{\mu}.$$

In the M/M/1/∞-system with (r, Q)- or (r, S)-policy according to Definition 3.1, resp. 4.1, we obtain

$$\alpha_{S,1} = 1 - \frac{\lambda}{\mu(\mu - (\lambda+\nu))}\left((\mu-\lambda)\left(\frac{\lambda}{\lambda+\nu}\right)^r - \nu\left(\frac{\lambda}{\mu}\right)^r\right),$$

and

$$s = r - \frac{\lambda}{\nu}\left(1 - \left(\frac{\lambda}{\lambda+\nu}\right)^r\right) + \frac{(1-\alpha_{S,1})\mu - \lambda}{\mu - \lambda}.$$

If we replace the first item in Corollary 5.3 by $1 - \frac{\lambda}{\mu} \leq \alpha_{S,1} \leq 1 - \left(\frac{\lambda}{\mu}\right)^r$ then Corollary 5.3 and 5.4 hold with similar proofs.

6 M/M/1/N-1-systems with different inventory management policies and lost sales

Let us now consider the systems from Sections 2, 3 and 4 with a limited number of N customers which can be admitted to the system. We assume that during the time a replenishment order is outstanding the service place can be used as a waiting place by the customers in the system.

6.1 M/M/1/N-1-system with inventory management, reorder point 0 and arbitrary size of replenishment orders

Definition 6.1 *An M/M/1/N-1-system with inventory management, reorder point 0 and arbitrary size of replenishment orders is a single server under FCFS regime as described in Definition 2.1 with a limited number of $N-1$ waiting places. If at the moment of arrival of a customer the number of customers present at the system is less than N and the on-hand inventory is positive, the arriving customer is admitted to the system, otherwise the arriving customer is not admitted to*

the system. It disappears and never returns.
Let $X(t)$ denote the number of customers present at the server at time $t \geq 0$, either waiting or in service (queue length) and let $Y(t)$ denote the on-hand inventory at time $t \geq 0$. Then $Z = ((X(t), Y(t)), t \geq 0)$ is a continuous-time Markov process for the M/M/1/N-1-system with inventory management, reorder point 0 and arbitrary size of replenishment orders. The state space of Z is $E_{Z_N} = \{(n,k) : n \in \{0,1,...,N\}, k \in \mathbb{N}\} \cup \{(n,0) : n \in \{0,1,...,N-1\}\}$, since the inventory can only be depleted after a customer has been served who took the last item from inventory and no customers join the queue during the replenishment lead time.

Theorem 6.2 *The continuous-time Markov process Z from Definition 6.1 is ergodic if and only if $\overline{p} < \infty$. If Z is ergodic then it has a unique limiting and stationary distribution of product form given by*

$$\pi(n,k) = K^{-1}\left(\frac{\lambda}{\mu}\right)^n q_k, \qquad \text{with } 0 \leq n \leq N,\ k \geq 1, \tag{43}$$

$$\pi(n,0) = K^{-1}\left(\frac{\lambda}{\mu}\right)^n \frac{\lambda}{\nu}, \qquad \text{with } 0 \leq n \leq N-1, \tag{44}$$

and with normalization constant

$$K = \frac{\mu^{N+1}\left(\overline{p} + \frac{\lambda}{\nu}\right) - \lambda^{N+1}\left(\overline{p} + \frac{\mu}{\nu}\right)}{\mu^N(\mu - \lambda)}. \tag{45}$$

Proof. The process Z from Definition 6.1 possesses the following global balance equations:

$$\begin{aligned}\pi(n,k)(\lambda(1-\delta_{Nn}) + \mu(1-\delta_{0n})) &= \pi(n-1,k)\lambda(1-\delta_{0n}) \\ &+\pi(n+1,k+1)\mu(1-\delta_{Nn}) + \pi(n,0)\nu p_k(1-\delta_{Nn}), \quad 0 \leq n \leq N, 1 \leq k,\end{aligned} \tag{46}$$

$$\pi(n,0)\nu = \pi(n+1,1)\mu, \quad 0 \leq n \leq N-1. \tag{47}$$

We have to show that the distribution from (43) and (44) satisfies equations (46) and (47). This is done by insertion using $q_1 = 1$ and $q_k = q_{k+1} + p_k$. The normalization constant K is given by

$$\begin{aligned} K &= \sum_{n=0}^{N}\sum_{k\geq 1}\left(\frac{\lambda}{\mu}\right)^n q_k + \sum_{n=0}^{N-1}\left(\frac{\lambda}{\mu}\right)^{n+1}\frac{\mu}{\nu} = \frac{1-\left(\frac{\lambda}{\mu}\right)^{N+1}}{1-\frac{\lambda}{\mu}}\left(\overline{p} + \frac{\mu}{\nu}\right) - \frac{\mu}{\nu} \\ &= \frac{\mu^{N+1}\left(\overline{p} + \frac{\lambda}{\nu}\right) - \lambda^{N+1}\left(\overline{p} + \frac{\mu}{\nu}\right)}{\mu^N(\mu - \lambda)}, \end{aligned}$$

where $\overline{p}$ denotes the mean size of a replenishment order. □

Remark 6.3 *The stationary distribution is the conditional distribution of the open system with infinite waiting room from Definition 2.1 and Theorem 2.2 conditioned on the total population size excluding the state $(N,0)$. Our pair of systems with finite resp. infinite waiting room shares this property with standard birth and death queues.*

We define

$$\begin{aligned} K_X &:= \frac{\mu^N - \lambda^N}{\mu^{N-1}(\mu - \lambda)} \\ K_Y &:= \frac{\mu^{N+1}\left(\overline{p} + \frac{\lambda}{\nu}\right) - \lambda^{N+1}\left(\overline{p} + \frac{\mu}{\nu}\right)}{\mu(\mu^N - \lambda^N)} \end{aligned}$$

such that $K = K_X \cdot K_Y$. Note that in case of a finite waiting room K cannot be splitted up in factors which only depend on queue length and inventory size, respectively.

The marginal steady state queue length probabilities are given as follows. For $n = 0, ..., N-1$

$$\begin{aligned}
\mathsf{P}(X = n) &= \sum_{k \geq 1} \pi(n, k) + \pi(n, 0) = K^{-1} \left(\frac{\lambda}{\mu}\right)^n \left(\overline{p} + \frac{\lambda}{\nu}\right), \\
\mathsf{P}(X = N) &= \sum_{k \geq 1} \pi(N, k) = K^{-1} \left(\frac{\lambda}{\mu}\right)^N \overline{p}.
\end{aligned}$$

These probabilities do not coincide with the steady state probabilities of the queue length process in the standard M/M/1/N-1-FCFS-system. The steady state probability that the on-hand inventory is $k \geq 1$ is given by

$$\mathsf{P}(Y = k) = \sum_{n=0}^{N} \pi(n, k) = q_k \frac{\mu^{N+1} - \lambda^{N+1}}{\mu^N(\mu - \lambda)} K^{-1}.$$

The probability that the server is depleted is given by

$$\mathsf{P}(Y = 0) = \sum_{n=0}^{N-1} \pi(n, 0) = \frac{\lambda}{\nu} \frac{\mu^N - \lambda^N}{\mu^{N-1}(\mu - \lambda)} K^{-1} = \frac{\lambda}{\nu} K_Y^{-1}.$$

Remark 6.4 *If $p_k = \delta_{k\infty}$ the system results in the standard M/M/1/N-1-system. If $\nu = \infty$ the inventory on hand leaves state 0 immediately and the state space is then given by $E_{Z_N} = \{(n, k) : n \in \{0, 1, ..., N\}, k \in \mathbb{N}\}$. If in addition the order size is fixed to some $Q \in \mathbb{N}$ then the inventory positions are uniformly distributed as in [BS00].*

Theorem 6.5 (Measures of system performance) *The measures of system performance as defined in Section 2.1 of the M/M/1/N-1-system with inventory management, reorder point 0 and arbitrary size of replenishment orders are given by*

$$\begin{aligned}
\overline{I} &= \frac{\mu^{N+1} - \lambda^{N+1}}{\mu^{N+1}\left(\overline{p} + \frac{\lambda}{\nu}\right) - \lambda^{N+1}\left(\overline{p} + \frac{\mu}{\nu}\right)} \sum_{k \geq 1} k q_k = \frac{\mu^{N+1} - \lambda^{N+1}}{K_Y \mu(\mu^N - \lambda^N)} \sum_{k \geq 1} k q_k, && (48) \\
\lambda_R &= \frac{(\mu^N - \lambda^N)\lambda\mu}{\mu^{N+1}\left(\overline{p} + \frac{\lambda}{\nu}\right) - \lambda^{N+1}\left(\overline{p} + \frac{\mu}{\nu}\right)} = \frac{\lambda}{K_Y}, && (49) \\
\lambda_A &= \frac{(\mu^N - \lambda^N)\lambda\mu\overline{p}}{\mu^{N+1}\left(\overline{p} + \frac{\lambda}{\nu}\right) - \lambda^{N+1}\left(\overline{p} + \frac{\mu}{\nu}\right)} = \frac{\lambda\overline{p}}{K_Y}, && (50) \\
\overline{LS} &= \lambda\frac{\lambda}{\nu} K_Y^{-1} + \lambda \left(\frac{\lambda}{\mu}\right)^N \overline{p} K^{-1}, && (51) \\
\overline{LS}_c &= \frac{\lambda}{\nu} + \left(\frac{\lambda}{\mu}\right)^N \overline{p} K_X^{-1}, && (52) \\
s &= 0 \quad \textit{for } Y_{net} = Y \textit{ and } Y_{net} = \max\{0, Y - X\}, && (53) \\
\alpha_1 &= 0 \quad \textit{for } Y_{net} = Y \textit{ and } Y_{net} = \max\{0, Y - X\}, && (54) \\
\alpha_2 &= \frac{(\mu^{N+1} - \lambda^{N+1})\overline{p}}{\mu^{N+1}\left(\overline{p} + \frac{\lambda}{\nu}\right) - \lambda^{N+1}\left(\overline{p} + \frac{\mu}{\nu}\right)} = \frac{\overline{p}}{K_Y} \frac{\mu^{N+1} - \lambda^{N+1}}{\mu(\mu^N - \lambda^N)} \quad \textit{for } Y_{net} = Y, && (55)
\end{aligned}$$

$$\beta = \frac{(\mu^N - \lambda^N)\overline{p}\mu}{\mu^{N+1}\left(\overline{p} + \frac{\lambda}{\nu}\right) - \lambda^{N+1}\left(\overline{p} + \frac{\mu}{\nu}\right)} = \frac{\overline{p}}{K_Y}, \tag{56}$$

$$\overline{W}_0 = \frac{1}{\mu - \lambda}\left(1 + \frac{\lambda}{\nu\overline{p}}\right) - \frac{N\lambda^N}{\mu(\mu^N - \lambda^N)}\left(1 + \frac{\mu}{\nu\overline{p}}\right), \tag{57}$$

$$\overline{W} = \frac{\lambda(\mu^{N-1} - \lambda^{N-1})}{(\mu - \lambda)(\mu^N - \lambda^N)}\left(1 + \frac{\lambda}{\nu\overline{p}}\right) - (N-1)\frac{\lambda^N}{\mu(\mu^N - \lambda^N)}\left(1 + \frac{\mu}{\nu\overline{p}}\right). \tag{58}$$

Proof. The proof of Theorem 3.5 is presented in Appendix B on page 31. □

Note that the mean time between two orders

$$\frac{1}{\lambda_R} = \frac{\mu^{N+1}\left(\overline{p} + \frac{\lambda}{\nu}\right) - \lambda^{N+1}\left(\overline{p} + \frac{\mu}{\nu}\right)}{(\mu^N - \lambda^N)\lambda\mu} = \frac{\mu^{N+1} - \lambda^{N+1}}{(\mu^N - \lambda^N)\lambda}\frac{\overline{p}}{\mu} + \frac{1}{\nu}$$

is larger than in the M/M/1/∞-system as defined in Definition 2.1 if $\mu > \lambda$.
$\overline{LS}$ and $\overline{LS}_c$ are now composed of two parts: the first part represents the lost customers from an empty inventory and the second part represents the lost customers from a full waiting room. As before $\beta = \frac{\lambda_A}{\lambda}$ but $\beta \neq \alpha_2 = \mathsf{P}(Y > 0)$. The mean sojourn time $\overline{W}_{0,s}$ and the mean waiting time $\overline{W}_s$ of the standard M/M/1/N-1-system are given by (see [GH85] Section 2.4, p. 97)

$$\overline{W}_{0,s} = \frac{1}{\mu - \lambda} - \frac{N\lambda^N}{\mu(\mu^N - \lambda^N)} \quad \text{and} \quad \overline{W}_s = \frac{\lambda}{\mu(\mu - \lambda)} - \frac{N\lambda^N}{\mu(\mu^N - \lambda^N)}.$$

Hence,

$$\overline{W}_0 - \overline{W}_{0,s} = \frac{1}{\mu - \lambda}\frac{\lambda}{\nu\overline{p}} - \frac{N\lambda^N}{\mu(\mu^N - \lambda^N)}\frac{\mu}{\nu\overline{p}} = \frac{\mu^N - N\lambda^{N-1}\mu + (N-1)\lambda^N}{(\mu - \lambda)(\mu^N - \lambda^N)}\frac{\lambda}{\nu\overline{p}} > 0.$$

A similar computation for the mean waiting times yields

$$\overline{W} - \overline{W}_s = \frac{\mu^{N-1} - (N-1)\lambda^{N-2}\mu + (N-2)\lambda^{N-1}}{(\mu - \lambda)(\mu^N - \lambda^N)}\frac{\lambda^2}{\nu\overline{p}} > 0.$$

Remark 6.6 *If $p_k = \delta_{k\infty}$ or $\nu \to \infty$ the following performance measures are the same as in the standard M/M/1/N-1-system: $\lambda_R = \frac{\lambda}{\overline{p}}\mathsf{P}(X < N)$, λ_A, $\overline{LS} = \lambda\mathsf{P}(X = N)$, $\beta = \mathsf{P}(X < N)$, $\overline{W}_0$ and $\overline{W}$.*

Remark 6.7 *In the case of a finite waiting room all performance measures depend on the service rate μ and the parameters λ and ν. The influence of F_p is the same as in the case of an infinite waiting room.*

6.2 M/M/1/N-1-system with (r, Q)-policy

In this section we consider the finite version of the M/M/1/∞-system with (r, Q)-policy which has been introduced in Section 3.

Definition 6.8 *An M/M/1/N-1-system with (r, Q)-policy is a single server under FCFS regime as described in Definition 3.1 but now the number of waiting places is limited to $N - 1$.
Let $X(t)$ denote the number of customers present at the server at time $t \geq 0$, either waiting or in service and let $Y(t)$ denote the on-hand inventory at time $t \geq 0$. Then $Z = ((X(t), Y(t)), t \geq 0)$ is a continuous-time Markov process for the M/M/1/N-1-system with (r, S)-policy. The state space of Z is $E_{Z_N} = \{(n, k) : n \in \{0, 1, ..., N\}, 1 \leq k \leq Q + r\} \cup \{(n, 0) : n \in \{0, 1, ..., N - 1\}\}$.*

Theorem 6.9 *If we assume that a replenishment order is outstanding only if the on-hand inventory is smaller or equal than r and if there are less than N customers present in the system, the continuous-time Markov process Z has a stationary distribution of product form given by*

$$\pi(n,k) = K^{-1}\left(\frac{\lambda}{\mu}\right)^n C(k), \quad \textit{with } 0 \le n \le N, 1 \le k \le r+Q, \tag{59}$$

$$\pi(n,0) = K^{-1}\left(\frac{\lambda}{\mu}\right)^n \frac{\lambda}{\nu}, \qquad \textit{with } 0 \le n \le N-1. \tag{60}$$

$C(k), 1 \le k \le Q+r$ is the same as in Theorem 3.3 and the normalization constant is given by

$$K = \frac{\mu^{N+1}-\lambda^{N+1}}{\mu^N(\mu-\lambda)}\left(Q\left(\frac{\lambda+\nu}{\lambda}\right)^r + \frac{\lambda}{\nu}\right) - \frac{\lambda}{\nu}\frac{\lambda^N}{\mu^N}. \tag{61}$$

Proof. The proof of Theorem 3.3 is modified by replacing ν by $\nu(1-\delta_{Nn})$. Furthermore, λ on the left side of (12) and μ on the right side of (12) and (13) are multiplied by $(1-\delta_{Nn})$. ☐

6.3 M/M/1/N-1-system with (r,S)-policy

In this section we consider the finite version of the M/M/1/∞-system with (r,S)-policy which has been introduced in Section 4.

Definition 6.10 *An M/M/1/N-1-system with (r,S)-policy is a single server under FCFS regime as described in Definition 4.1 but now the number of waiting places is limited to $N-1$.*
Let $X(t)$ denote the number of customers present at the server at time $t \ge 0$, either waiting or in service and let $Y(t)$ denote the on-hand inventory at time $t \ge 0$. Then $Z = ((X(t),Y(t)), t \ge 0)$ is a continuous-time Markov process for the M/M/1/N-1-system with (r,S)-policy. The state space of Z is $E_{Z_N} = \{(n,k) : n \in \{0,1,...,N\}, 1 \le k \le S\} \cup \{(n,0) : n \in \{0,1,...,N-1\}\}$.

Theorem 6.11 *If we assume that a replenishment order is outstanding only if the inventory on hand is smaller or equal than r and if there are less than N customers present in the system, the continuous-time Markov process Z has a stationary distribution of product form given by*

$$\pi(n,k) = K^{-1}\left(\frac{\lambda}{\mu}\right)^n C(k), \quad \textit{with } 0 \le n \le N, 1 \le k \le S, \tag{62}$$

$$\pi(n,0) = K^{-1}\left(\frac{\lambda}{\mu}\right)^n \frac{\lambda}{\nu}, \qquad \textit{with } 0 \le n \le N-1. \tag{63}$$

$C(k), 1 \le k \le S$ is the same as in Theorem 4.3 and the normalization constant is given by

$$K = \frac{\mu^{N+1}-\lambda^{N+1}}{\mu^N(\mu-\lambda)}\left(S-r+\frac{\lambda}{\nu}\right)\left(\frac{\lambda+\nu}{\lambda}\right)^r - \frac{\lambda}{\nu}\frac{\lambda^N}{\mu^N}. \tag{64}$$

Proof: The proof of Theorem 4.3 is modified by replacing ν by $\nu(1-\delta_{Nn})$. Furthermore, λ on the left side of (29) and μ on the right side of (29) and (30) are multiplied by $(1-\delta_{Nn})$. ☐

7 Conclusions

In this paper we have studied M/M/1 queueing systems with attached inventory, exponentially distributed lead times and lost sales for infinite and finite waiting rooms. The discussed inventory management policies are the (r, Q)- and (r, S)-policies and a policy with reorder point 0 and arbitrary size of replenishment orders. For each of these systems we have determined the stationary distribution. Surprisingly enough all these distributions are of product form. In case of infinite waiting room the limiting distributions of the queue length processes coincide in addition with that of the M/M/1/∞-system. To the best of our knowledge these are the first closed form solutions to service facilities with attached inventories and stochastic lead times. Note that state dependent service rates can be incorporated into the different models without any difficulty.
Various definitions for performance measures found in the literature have been quoted and compared in the underlying models.
Directions for future research are an investigation of the long-run average cost functions with/without a service constraint and to allow for more general lead time and service time distributions.

A Customer and event stationary measures and intensities

This section contains results on intensities for the customer arrival/departure processes and the replenishment order arrival process for the systems from Sections 2, 3, 4 and 6. Besides these performance measures the Palm probability associated with the point process of replenishments is given. These probabilities are needed to calculate the safety stock and α_1-service level as defined in Definitions 2.7 and 2.8.
In addition to these results used in earlier sections we also point out some results related to the arrival theorem for systems with limited and unlimited population size.

We assume that under P Z is a regular, stationary Markov Process with Q-Matrix $Q = \{q(i,j)\}$ and invariant measure π.

Theorem A.1 *In the M/M/1/∞-system with inventory management, reorder point 0 and arbitrary size of replenishment orders according to Definition 2.1 the intensity of departures λ_D and the intensity of arrivals of customers who are admitted to the system λ_A are given by*

$$\lambda_D = \lambda_A = \frac{\overline{p}\lambda\nu}{\overline{p}\nu + \lambda}. \tag{65}$$

The intensity of arrivals of replenishment orders λ_R is given by

$$\lambda_R = \frac{\lambda\nu}{\overline{p}\nu + \lambda}. \tag{66}$$

The Palm probability associated with the point process of replenishments is

$$\mathsf{P}^0_{N_R}(Z_{0-} = (n,0)) \quad = \quad \frac{\overline{p}\nu + \lambda}{\lambda}\,\pi(n,0) \qquad\qquad \text{for } n \in \mathbb{N}_0. \tag{67}$$

Proof. Let H be a subset of $E_Z \times E_Z -$ diag $(E_Z \times E_Z)$ and let N_H be the point process counting the H-transitions of Z, i.e.

$$N_H(C) = \int_C 1_H(Z_{s-}, Z_s)\, N(ds),\; C \in \mathbb{B}.$$

Let us consider special sets that describe events of the queueing system under investigation

$$
\begin{array}{lll}
D = & \{\big((n,k),(n-1,k-1)\big) : n \geq 1, k \geq 1\} & \text{departure of customers,} \\
A = & \{\big((n,k),(n+1,k)\big) : n \geq 0, k \geq 1\} & \text{arrival of customers,} \\
 & & \text{who are admitted to the system,} \\
R_k = & \{\big((n,0),(n,k)\big) : n \geq 0\} & \text{arrival of replenishment} \\
 & & \text{orders of size } k \geq 1, \\
R = & \{\big((n,0),(n,k)\big) : n \geq 0, k \geq 1\} & \text{arrival of replenishment} \\
 & & \text{orders.}
\end{array}
$$

From formula (1.4.15) on page 38 of [BB03]

$$
\lambda_H = E[N_H((0,1])] = \sum_{((n,k),(m,l)) \in H} \pi(n,k) q\big((n,k),(m,l)\big) \tag{68}
$$

and therefore

$$
\begin{aligned}
\lambda_D &= \sum_{((n,k),(n-1,k-1)) \in D} \pi(n,k) q\big((n,k),(n-1,k-1)\big) \\
&= \mu \sum_{n \geq 1} \sum_{k \geq 1} \pi(n,k) = \mu \frac{\lambda}{\mu} \frac{\overline{p}}{\overline{p} + \frac{\lambda}{\nu}} = \frac{\overline{p}\lambda\nu}{\overline{p}\nu + \lambda} = \frac{\lambda \overline{p}}{K_Y}, \\
\lambda_A &= \sum_{((n,k),(n+1,k)) \in A} \pi(n,k) q\big((n,k),(n+1,k)\big) = \lambda \mathsf{P}(Y > 0) = \lambda_D, \\
\lambda_{R_k} &= \sum_{((n,0),(n,k)) \in R_k} \pi(n,0) q\big((n,0),(n,k)\big) = \nu p_k \mathsf{P}(Y = 0) = \frac{\lambda p_k}{K_Y}, \\
\lambda_R &= \sum_{k \geq 1} \lambda_{R_k} = \frac{\lambda}{K_Y}.
\end{aligned}
$$

Since $0 < \lambda_H < \infty$ we can define the Palm probability $P^0_{N_H}$ associated with N_H for all $H \in \{D, A, R_k, R\}$ as in formulas (1.4.18) and (1.4.19) on page 39 of [BB03].

$$
\begin{aligned}
\mathsf{P}^0_{N_D}(Z_0 = (n,k)) &= \frac{\sum_{(m,l)/((m,l),(n,k)) \in D} \pi(m,l) q((m,l),(n,k))}{\lambda_D} \\
&= \frac{\mu(\overline{p}\nu + \lambda)}{\overline{p}\lambda\nu} \pi(n+1,k+1) = \frac{K_Y}{\overline{p}} \pi(n,k+1), n,k \in \mathbb{N}_0 \\
\mathsf{P}^0_{N_D}(Z_{0-} = (n,k)) &= \begin{cases} \dfrac{\mu(\overline{p}\nu + \lambda)}{\overline{p}\lambda\nu} \pi(n,k) & \text{for } 0 < n,\ 0 < k, \\ 0 & \text{otherwise,} \end{cases} \\
&= \begin{cases} \dfrac{K_Y}{\overline{p}} \pi(n-1,k) & \text{for } 0 < n,\ 0 < k, \\ 0 & \text{otherwise,} \end{cases} \\
\mathsf{P}^0_{N_A}(Z_0 = (n,k)) &= \frac{\sum_{(m,l)/((m,l),(n,k)) \in A} \pi(m,l) q((m,l),(n,k))}{\lambda_A} \\
&= \begin{cases} \dfrac{K_Y}{\overline{p}} \pi(n-1,k) & \text{for } 0 < n, 0 < k, \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{P}^0_{N_A}(Z_{0-}=(n,k)) &= \pi(n,k)\frac{\sum_{(m,l)/((n,k),(m,l))\in A} q((n,k),(m,l))}{\lambda_A} \\
&= \begin{cases} \dfrac{K_Y}{\overline{p}}\pi(n,k) & \text{for } n\in\mathbb{N}_0\,, 0<k, \\ 0 & \text{otherwise,} \end{cases} \\
\mathsf{P}^0_{N_{R_k}}(Z_0=(n,k)) &= \frac{\sum_{(m,l)/((m,l),(n,k))\in R_k} \pi(m,l)q((m,l),(n,k))}{\lambda_{R_k}} \\
&= \frac{K_Y\nu}{\lambda}\pi(n,0) \qquad \text{for } n\in\mathbb{N}_0\,, \\
\mathsf{P}^0_{N_R}(Z_{0-}=(n,0)) &= \pi(n,0)\frac{\sum_{(m,l)/((n,0),(m,l))\in R} q((n,0),(m,l))}{\lambda_R} \\
&= \frac{K_Y\nu}{\lambda}\pi(n,0) \qquad \text{for } n\in\mathbb{N}_0\,. \qquad \square
\end{aligned}
$$

For a large class of networks having product-form equilibrium distribution it is known (see [SM81]) that if a customer belongs to an open subchain the state distribution at arrival points, departure points, and random points are identical. In this system the arrival-instant/departure-instant and the steady-state distributions are not the same. Just before an arrival a customer observes

$$\mathsf{P}^0_{N_A}(Z_{0-}=(n,k)) = \frac{K_Y}{\overline{p}}\pi(n,k), \quad n\in\mathbb{N}_0\,, 0<k$$

whereas just after a departure he leaves

$$\mathsf{P}^0_{N_D}(Z_0=(n,k)) = \frac{K_Y}{\overline{p}}\pi(n,k+1) \quad n\in\mathbb{N}_0\,, 0<k.$$

These probabilities are not equal since $\pi(n,k)\neq\pi(n,k+1)$. Note that an arriving customer never sees the state $(n,0)$ without counting himself, whereas a departing customer can leave the state $(n,0)$ if he gets the last item from inventory. The following Palm probabilities are equal.

$$\mathsf{P}^0_{N_A}(Z_0=(n,k)) = \mathsf{P}^0_{N_D}(Z_{0-}=(n,k)), \quad 0<n, 0<k.$$

After an arrival a customer observes the state (n,k) with the same probability as a customer just before his departure.

If we look at aggregate states $n\in\mathbb{N}_0$, defined as the queue length, we obtain

$$
\begin{aligned}
\mathsf{P}^0_{N_A}(X_{0-}=n) &= \frac{K_Y}{\overline{p}}\sum_{k\geq 1}\pi(n,k) \\
&= K_X\left(\frac{\lambda}{\mu}\right)^n = \mathsf{P}(X=n) \\
&= \frac{K_Y}{\overline{p}}\sum_{k\geq 0}\pi(n,k+1) = \mathsf{P}^0_{N_D}(X_0=n).
\end{aligned}
$$

This shows that the distributions of the queue length $n\in\mathbb{N}_0$ just before an arrival and just after a departure instant and in the time stationary queue are the same. This is the ESTA (events see time averages) property [BvKM92], note that arrivals to the server do not form a Poisson process. For $0<n$ we compute

$$\mathsf{P}^0_{N_A}(X_0=n) = \frac{K_Y}{\overline{p}}\sum_{k>0}\pi(n-1,k) = \mathsf{P}(X=n-1) = \mathsf{P}^0_{N_D}(X_{0-}=n).$$

Hence, the distributions of n customers in the system just after an arrival and just before a departure instant are the same and they are equal to the time stationary probability with one less customer in the queue. As for every stochastic process on $\mathbb{N}_0$ with step size one, the steady state probabilities in arrival and departure instants are equal if they exist.

Theorem A.2 *In the M/M/1/∞-system with (r, Q)-policy according to Definition 3.1 the intensity of departures λ_D and the intensity of arrivals of customers who are admitted to the system λ_A are given by*

$$\lambda_D = \lambda_A = \frac{Q\lambda}{Q + \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}. \tag{69}$$

The intensity of arrivals of replenishment orders λ_R is given by

$$\lambda_R = \frac{\lambda}{Q + \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}. \tag{70}$$

The Palm probability associated with the point process of replenishments is

$$\mathsf{P}^0_{N_R}(Z_{0-} = (n,h)) \quad = \quad \frac{\nu}{\lambda_R}\pi(n,h) \qquad \textit{for } 0 \le h \le r, n \in \mathbb{N}_0. \tag{71}$$

Proof. Define the special sets D for the departure of customers, A for the arrival of customers, who are admitted to the system, and R for arrivals of replenishment orders as before. Using equation (68) it is straight forward to prove that in the context of the M/M/1/∞-system with (r, Q)-policy the intensities λ_D, λ_A and λ_R are equal to the expressions in (69) and (70). For $0 \le h \le r$ and $n \in \mathbb{N}_0$

$$\mathsf{P}^0_{N_R}(Z_{0-} = (n,h)) \quad = \quad \pi(n,h)\frac{\sum_{(m,l)/((n,h),(m,l))\in R} q((n,h),(m,l))}{\lambda_R} = \frac{\nu}{\lambda_R}\pi(n,h). \qquad \square$$

Theorem A.3 *In the M/M/1/∞-system with (r, S)-policy according to Definition 4.1 the intensity of departures λ_D and the intensity of arrivals of customers who are admitted to the system λ_A are given by*

$$\lambda_D = \lambda_A = \lambda - \frac{\lambda^2}{(S-r)\nu + \lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r. \tag{72}$$

The intensity of arrivals of replenishment orders λ_R is given by

$$\lambda_R = \frac{\lambda}{S - r + \frac{\lambda}{\nu}}. \tag{73}$$

The Palm probability associated with the point process of replenishments is

$$\mathsf{P}^0_{N_R}(Z_{0-} = (n,h)) \quad = \quad \frac{\nu}{\lambda_R}\pi(n,h) \qquad \textit{for } 0 \le h \le r, n \in \mathbb{N}_0. \tag{74}$$

Proof. The proof can be done analogously to the proof of Theorem A.2. $\square$

The observations made above concerning the probabilities in arrival and departure instants carry over to the M/M/1/∞-system with (r, Q)-policy or (r, S)-policy.

Theorem A.4 *In the M/M/1/N-1-system with inventory management, reorder point* 0 *and arbitrary size of replenishment orders according to Definition 6.1 the intensity of departures* λ_D *and the intensity of arrivals of customers who are admitted to the system* λ_A *are given by*

$$\lambda_D = \lambda_A = \frac{\overline{p}\lambda}{K_Y}. \tag{75}$$

The intensity of arrivals of replenishment orders λ_R *is given by*

$$\lambda_R = \frac{\lambda}{K_Y}. \tag{76}$$

The Palm probability associated with the point process of replenishments is

$$\mathsf{P}^0_{N_R}(Z_{0-} = (n,0)) \quad = \quad \frac{K_Y \nu}{\lambda}\pi(n,0) \qquad\qquad \text{for } 0 \leq n \leq N-1. \tag{77}$$

Proof. As before we consider the following special sets that describe events of the queueing system

$D =$	$\{\big((n,k),(n-1,k-1)\big) : 1 \leq n \leq N, 1 \leq k\}$	departure of customers,
$A =$	$\{\big((n,k),(n+1,k)\big) : 0 \leq n \leq N-1, 1 \leq k\}$	arrival of customers,
$R_k =$	$\{\big((n,0),(n,k)\big) : 0 \leq n \leq N-1\}$	arrival of a replenishment order of size $k \geq 1$,
$R =$	$\{\big((n,0),(n,k)\big) : 0 \leq n \leq N-1, 1 \leq k\}$	arrival of a replenishment order.

Computations similar to those in the proof of Theorem A.1 show that λ_D, λ_A and λ_R have the form given in (75) and (76). The calculation of $\mathsf{P}^0_{N_R}(Z_{0-} = (n,0))$ is straight forward. $\square$

We will now study the finite queueing system at the moments when customers arrive at the server or leave the server. In this system the arrival-instant/departure-instant and the steady-state distributions are not the same. We have

$$\mathsf{P}^0_{N_A}(Z_{0-} = (n,k)) \neq \mathsf{P}^0_{N_D}(Z_0 = (n,k)), \quad 0 < k, 0 \leq n \leq N-1$$

since $\pi(n,k) \neq \pi(n,k+1)$ and

$$\mathsf{P}^0_{N_A}(Z_0 = (n,k)) = \mathsf{P}^0_{N_D}(Z_{0-} = (n,k)), \quad 0 < n \leq N, 0 < k.$$

If we look at aggregate states n for $0 \leq n \leq N-1$, we obtain

$$\mathsf{P}^0_{N_A}(X_{0-} = n) \quad = \quad \mathsf{P}^0_{N_D}(X_0 = n) = \frac{K_Y}{\overline{p}} \sum_{k \geq 1} \pi(n,k) = K_{ql}^{-1}\left(\frac{\lambda}{\mu}\right)^n \neq \mathsf{P}(X = n).$$

This shows that the distributions of the queue length n just before an arrival and just after a departure instant are the same. The time stationary distribution of the queue length is different. For $0 < n$ we compute

$$\mathsf{P}^0_{N_A}(X_0 = n) = \mathsf{P}^0_{N_D}(X_{0-} = n) = \frac{K_Y}{\overline{p}} \sum_{k>0} \pi(n-1,k) \neq \mathsf{P}(X = n-1). \tag{78}$$

Hence, the distributions of n customers in the system just after an arrival or just before a departure instant are not equal to the time stationary probability with one less customer in the queue. This is in contrast to the arrival theorem [SM81] for networks with limited population size.

B In the main text omitted proofs

Proof of Theorem 3.5 The average on-hand inventory position is given by

$$
\begin{aligned}
\overline{I} &= \sum_{k=1}^{Q+r} k \sum_{n\geq 0} \pi(n,k) \\
&= \left(Q\left(\frac{\lambda+\nu}{\lambda}\right)^r + \frac{\lambda}{\nu}\right)^{-1} \left\{ \sum_{k=1}^{r} k \left(\frac{\lambda+\nu}{\lambda}\right)^{k-1} + \sum_{k=r+1}^{Q} k \left(\frac{\lambda+\nu}{\lambda}\right)^r \right. \\
&\qquad \left. + \sum_{k=1}^{r} (Q+k) \left[\left(\frac{\lambda+\nu}{\lambda}\right)^r - \left(\frac{\lambda+\nu}{\lambda}\right)^{k-1} \right] \right\} \\
&= \left(Q\left(\frac{\lambda+\nu}{\lambda}\right)^r + \frac{\lambda}{\nu}\right)^{-1} \left\{ -\sum_{k=1}^{r} Q \left(\frac{\lambda+\nu}{\lambda}\right)^{k-1} + \sum_{k=r+1}^{Q+r} k \left(\frac{\lambda+\nu}{\lambda}\right)^r \right\} \\
&= \left(Q\left(\frac{\lambda+\nu}{\lambda}\right)^r + \frac{\lambda}{\nu}\right)^{-1} \left\{ \frac{Q}{\lambda^{r-1}} \frac{\lambda^r - (\lambda+\nu)^r}{\nu} \right. \\
&\qquad \left. + \left(\frac{\lambda+\nu}{\lambda}\right)^r \left[\frac{(Q+r+1)(Q+r)}{2} - \frac{(r+1)r}{2} \right] \right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{Q}{Q\left(\frac{\lambda+\nu}{\lambda}\right)^r + \frac{\lambda}{\nu}} \left\{ \frac{\lambda}{\nu} - \frac{\lambda}{\nu}\left(\frac{\lambda+\nu}{\lambda}\right)^r + \left(\frac{\lambda+\nu}{\lambda}\right)^r \left[\frac{Q+1}{2} + r\right] \right\} \\
&= \frac{Q}{Q\left(\frac{\lambda+\nu}{\lambda}\right)^r + \frac{\lambda}{\nu}} \left\{ \left(\frac{Q+1}{2} + r - \frac{\lambda}{\nu}\right) \left(\frac{\lambda+\nu}{\lambda}\right)^r + \frac{\lambda}{\nu} \right\} \\
&= \frac{Q}{Q + \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r} \left\{ \frac{Q+1}{2} + r - \frac{\lambda}{\nu} + \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r \right\}.
\end{aligned}
$$

Using s from (20), that will be calculated below, $\overline{I}$ is obtained in the form given in (15). λ_R and λ_A are a result of Theorem A.2.

$$
\begin{aligned}
\overline{LS} &= \lambda \sum_{n\geq 0} \pi(n,0) = \lambda \mathsf{P}(Y=0) = \lambda \frac{\frac{\lambda}{\nu}}{Q\left(\frac{\lambda+\nu}{\lambda}\right)^r + \frac{\lambda}{\nu}}, \\
\overline{LS}_c &= \frac{\overline{LS}}{\lambda_R} = \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r.
\end{aligned}
$$

Involving Definition 2.7 and (71) from Theorem A.2 the safety stock or the expected net inventory position just before a replenishment order arrives is

$$
\begin{aligned}
s &= \sum_{k\geq 1} k\, \mathsf{P}^0_{N_R}(Y_{net,-} = k) = \sum_{k=1}^{r} k \sum_{n\geq 0} \mathsf{P}^0_{N_R}(Z_{0-} = (n,k)) \\
&= \frac{\nu}{\lambda_R} \sum_{k=1}^{r} k \sum_{n\geq 0} \pi(n,k) = \frac{\nu}{\lambda_R} K_Y^{-1} \sum_{k=1}^{r} k C(k) \\
&= \frac{\nu}{\lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r \frac{1-(r+1)\left(\frac{\lambda+\nu}{\lambda}\right)^r + r\left(\frac{\lambda+\nu}{\lambda}\right)^{r+1}}{\left(1-\frac{\lambda+\nu}{\lambda}\right)^2}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\nu}{(\lambda+\nu)^r}\frac{\lambda^{r+1}-(r+1)\lambda(\lambda+\nu)^r+r(\lambda+\nu)^{r+1}}{\nu^2}\\
&= r-\frac{\lambda}{\nu}\left(1-\left(\frac{\lambda}{\lambda+\nu}\right)^r\right).
\end{aligned}
$$

Using Definitions 2.7 and 2.8 we have

$$
\begin{aligned}
\alpha_1 &= \mathsf{P}^0_{N_R}(Y_{0-}>0)=1-\sum_{n\geq 0}\mathsf{P}^0_{N_R}(Z_{0-}=(n,0))\\
&= 1-\frac{\nu}{\lambda_R}K_Y^{-1}\frac{\lambda}{\nu}=1-\left(\frac{\lambda}{\lambda+\nu}\right)^r.
\end{aligned}
$$

From Definition 2.8 $\alpha_2=\mathsf{P}(Y>0)$ and $\beta=(\lambda-\overline{LS})/\lambda$ hence, the results for α_2 and β are obvious. The mean number of customers in the system $\overline{L}_0$ and the mean number of waiting customers $\overline{L}$ are the same as in the usual M/M/1/∞-system. Using Little's formula the mean sojourn time $\overline{W}_0$ and the mean waiting time $\overline{W}$ of the customers are given by

$$
\begin{aligned}
\overline{W}_0 &= \frac{\overline{L}_0}{\lambda_A}=\frac{Q+\frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}{\lambda Q}\frac{\lambda}{\mu-\lambda}=\frac{1}{\mu-\lambda}+\frac{\lambda}{Q\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r\frac{1}{\mu-\lambda},\\
\overline{W} &= \frac{\overline{L}}{\lambda_A}=\frac{Q+\frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}{\lambda Q}\frac{\lambda^2}{\mu(\mu-\lambda)}=\frac{\lambda}{\mu(\mu-\lambda)}+\frac{\lambda}{Q\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r\frac{\lambda}{\mu(\mu-\lambda)}. \qquad \square
\end{aligned}
$$

Proof of Theorem 4.4 The average on-hand inventory position is given by

$$
\begin{aligned}
\overline{I} &= \sum_{k=1}^{S}k\sum_{n\geq 0}\pi(n,k)\\
&= \frac{1}{\left(S-r+\frac{\lambda}{\nu}\right)\left(\frac{\lambda+\nu}{\lambda}\right)^r}\left\{\sum_{k=1}^{r}k\left(\frac{\lambda+\nu}{\lambda}\right)^{k-1}+\sum_{k=r+1}^{S}k\left(\frac{\lambda+\nu}{\lambda}\right)^r\right\}\\
&= \frac{1}{\left(S-r+\frac{\lambda}{\nu}\right)\left(\frac{\lambda+\nu}{\lambda}\right)^r}\left\{\frac{1-(r+1)\left(\frac{\lambda+\nu}{\lambda}\right)^r+r\left(\frac{\lambda+\nu}{\lambda}\right)^{r+1}}{\left(1-\frac{\lambda+\nu}{\lambda}\right)^2}\right.\\
&\qquad \left.+\left(\frac{\lambda+\nu}{\lambda}\right)^r\frac{1}{2}\left((S+1)S-(r+1)r\right)\right\}\\
&= \frac{1}{S-r+\frac{\lambda}{\nu}}\left\{\frac{\lambda^2}{\nu^2}\left(\left(\frac{\lambda}{\lambda+\nu}\right)^r-1\right)+\frac{\lambda}{\nu}r+\frac{(S+1)S-(r+1)r}{2}\right\}\\
&= \frac{1}{S-r+\frac{\lambda}{\nu}}\left\{\frac{\lambda}{\nu}\left(r-\frac{\lambda}{\nu}+\frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r\right)+\frac{(S+1)S-(r+1)r}{2}\right\}.
\end{aligned}
$$

Using s from (37) the final form of $\overline{I}$ given in (32) follows. λ_R and λ_A are a result of Theorem A.3. The average number of lost sales incurred per unit of time and per cycle are given by

$$
\begin{aligned}
\overline{LS} &= \lambda\sum_{n\geq 0}\pi(n,0)=\lambda\mathsf{P}(Y=0)=\frac{\lambda^2}{(S-r)\nu+\lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r,\\
\overline{LS}_c &= \frac{\overline{LS}}{\lambda_R}=\frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r.
\end{aligned}
$$

Since $\lambda_R=\lambda\left(\frac{\lambda+\nu}{\lambda}\right)^r K_Y^{-1}$ as in the M/M/1/∞-system with (r,Q)-policy (see Theorems A.2 and A.3) and the constants $C(k)$ are the same in both systems for $1\leq k\leq r$ (see Theorems 3.3 and

4.3) the calculation of s in the proof of Theorem 3.5 carries over to the M/M/1/∞-system with (r,S)-policy. The same reasoning holds for the calculation of α_1. The β-service level is given by

$$\beta = \frac{\lambda - \overline{LS}}{\lambda} = 1 - \frac{\lambda}{(S-r)\nu + \lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r.$$

As before $\beta = \lambda_A/\lambda = \mathsf{P}(Y>0) = \alpha_2$.
The mean number of customers in the system $\overline{L}_0$ and the mean number of waiting customers $\overline{L}$ are the same as in the usual M/M/1/∞-system. Using Little's formula the formulas for the mean sojourn time $\overline{W}_0$ and the mean waiting time $\overline{W}$ can be verified to be of the form given in Theorem 4.4. □

Proof of Theorem 4.6 Let $S = Q + s$, where s is the safety stock in the M/M/1/∞-system with (r,Q)-policy from (20). Because of (37) $s = r - \frac{\lambda}{\nu}\left(1 - \left(\frac{\lambda}{\lambda+\nu}\right)^r\right)$ is also the safety stock of the order-up-to level (r,S)-system. In the (r,S)-system λ_R is given by

$$\lambda_R = \frac{\lambda}{S - s + \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}.$$

If $S - s = Q$ this is equal to λ_R in the (r,Q)-system. Using (37) and $S - s = Q$ once again we obtain from (34)

$$\lambda_A = \lambda - \frac{\lambda^2}{Q\nu + \lambda\left(\frac{\lambda}{\lambda+\nu}\right)^r}\left(\frac{\lambda}{\lambda+\nu}\right)^r = \frac{Q\lambda}{Q + \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}.$$

This is (17). Since $\overline{LS} = \lambda - \lambda_A$ and $\overline{LS}_c = \overline{LS}\lambda_R^{-1}$ the expected lost sales per unit of time and per cycle are the same in the (r,Q)- and (r,S)-system. α_1 is the same in both systems if the reorder point r is identical. From the definition $\beta = \frac{\lambda - \overline{LS}}{\lambda}$ coincides in the two systems if $\overline{LS}$ is the same. The equality of $\overline{W}_0$ and $\overline{W}$ can be seen from the formulas (23) and (40) and (24) and (41) respectively if $S - s = Q$ holds.

We will now show that if $S = Q + s$ the stationary mean inventory position $\overline{I}_l$ in the M/M/1/∞-system with order-level (r,S)-policy is not larger than the stationary mean inventory position $\overline{I}_q$ in the M/M/1/∞-system with order-quantity (r,Q)-policy. In the (r,Q)-system we have

$$\overline{I}_q = \frac{Q}{Q + \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}\left(\frac{Q+1}{2} + s\right).$$

Whereas in the (r,S)-system we have

$$\overline{I}_l = \frac{\frac{\lambda}{\nu}s + \frac{(S+1)S-(r+1)r}{2}}{S - s + \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r}.$$

Replacing Q by $S - s$ in the first expression we have

$$\overline{I}_l - \overline{I}_q = \frac{2\frac{\lambda}{\nu}s + (S+1)S - (r+1)r - (S-s)(S+s+1)}{2\left(S - s + \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r\right)} = \frac{2\frac{\lambda}{\nu}s - (r+1)r + (s+1)s}{2\left(S - s + \frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r\right)}.$$

Replacing the safety stock s by $r - \frac{\lambda}{\nu}\left(1 - \left(\frac{\lambda}{\lambda+\nu}\right)^r\right)$ (from (20) or (37)) and introducing $a := \frac{\lambda}{\lambda+\nu}$ and $b := \lambda/\nu$ yields

$$\begin{aligned}
\overline{I}_l - \overline{I}_q &= \frac{2b(r-b(1-a^r)) - (r+1)r + (r-b(1-a^r)+1)(r-b(1-a^r))}{2(S-r+b)} \\
&= b\frac{2(r-b(1-a^r)) - (r+1)(1-a^r) - (1-a^r)(r-b(1-a^r))}{2(S-r+b)} \\
&= b\frac{2r - (1-a^r)(b+2r+1+ba^r)}{2(S-r+b)} \\
&= b\frac{2ra^r - (1-a^r)(b+1+ba^r)}{2(S-r+b)} \\
&= b\frac{2ra^r - (1-a^r) - b(1-a^{2r})}{2(S-r+b)} \\
&\overset{\star}{=} b\frac{2ra^r(1-a) - (1-a^r)(1-a) - a(1-a^{2r})}{2(1-a)(S-r+b)} \\
&= b\frac{2ra^r(1-a) - (1-a^r)(1+a^{r+1})}{2(1-a)(S-r+b)} \\
&= \frac{b}{2(S-r+b)}\left(2ra^r - (1+a+a^2+\ldots+a^{r-1})(1+a^{r+1})\right) \\
&= \frac{b}{2(S-r+b)}\left(\sum_{k=0}^{r-1}(a^r-a^k) + \sum_{k=0}^{r-1}(a^r - a^{r+1+k})\right) \\
&= \frac{b}{2(S-r+b)}\left(\sum_{k=1}^{r} a^{r-k}(a^k-1) + \sum_{k=1}^{r} a^r(1-a^k)\right) \\
&= -\frac{b}{2(S-r+b)}\sum_{k=1}^{r} a^{r-k}(1-a^k)^2 < 0.
\end{aligned}$$

where $\overset{\star}{=}$ follows from $b = a/(1-a)$. Hence, $I_l < I_q$ and the proof is complete. □

Proof of Theorem 6.5 The average on-hand inventory is given by

$$\begin{aligned}
\overline{I} &= \sum_{k\geq 1} k \sum_{n=0}^{N} \pi(n,k) = K^{-1}\sum_{k\geq 1} kq_k \sum_{n=0}^{N}\pi(n,k) \\
&= \frac{\mu^{N+1} - \lambda^{N+1}}{\mu^{N+1}\left(\overline{p}+\frac{\lambda}{\nu}\right) - \lambda^{N+1}\left(\overline{p}+\frac{\mu}{\nu}\right)}\sum_{k\geq 1} kq(k).
\end{aligned}$$

λ_R and λ_A are given in Theorem A.4. The average number of lost sales incurred per unit of time and per cycle are given by

$$\begin{aligned}
\overline{LS} &= \lambda\left(\sum_{n=0}^{N-1}\pi(n,0) + \sum_{k=1}^{\infty}\pi(N,k)\right) = \lambda K^{-1}\left(\frac{\lambda}{\nu}\frac{\mu^N - \lambda^N}{\mu^{N-1}(\mu-\lambda)} + \left(\frac{\lambda}{\mu}\right)^N \overline{p}\right) \\
&= \lambda\frac{\lambda}{\nu}K_Y^{-1} + \lambda\left(\frac{\lambda}{\mu}\right)^N \overline{p}K^{-1}, \\
\overline{LS}_c &= \frac{\overline{LS}}{\lambda_R} = \frac{\lambda}{\nu} + \left(\frac{\lambda}{\mu}\right)^N \overline{p}K_X^{-1} = \frac{\lambda}{\nu} + \left(\frac{\lambda}{\mu}\right)^N \overline{p}K_X^{-1}.
\end{aligned}$$

Just as in the M/M/1/∞-system with inventory management, reorder point 0 and arbitrary size of replenishment orders treated in Section 2, the safety stock is 0 for both definitions of net inventory Y_{net}.
For the service levels we compute for the first definition of net inventory from 2.7

$$\alpha_1 = \mathsf{P}^0_{N_R}(Y_{net,0-} > 0) = \mathsf{P}^0_{N_R}(Y_{0-} > 0) = 0$$

and the second definition of net inventory from (42)

$$\alpha_1 = \mathsf{P}^0_{N_R}(Y_{net,0-} > 0) = \mathsf{P}^0_{N_R}(\max\{0, Y_{0-} - X_{0-}\} > 0) = 0.$$

$$\begin{aligned}
\alpha_2 &= \mathsf{P}(Y > 0) = 1 - \frac{\lambda}{\nu K_Y} = \frac{\mu^{N+1}(\overline{p} + \frac{\lambda}{\nu}) - \lambda^{N+1}(\overline{p} + \frac{\mu}{\nu}) - \frac{\lambda}{\nu}\mu(\mu^N - \lambda^N)}{\mu(\mu^N - \lambda^N)K_Y} \\
&= \frac{(\mu^{N+1} - \lambda^{N+1})\overline{p}}{\mu(\mu^N - \lambda^N)K_Y}. \\
\beta &= \frac{\lambda - \overline{LS}}{\lambda} = \frac{\overline{p}}{\overline{p} + \overline{LS}_c} = \frac{\overline{p}}{\overline{p}\frac{\mu^{N+1} - \lambda^{N+1}}{\mu(\mu^N - \lambda^N)} + \frac{\lambda}{\nu}} \\
&= \frac{\overline{p}\mu(\mu^N - \lambda^N)}{\overline{p}(\mu^{N+1} - \lambda^{N+1}) + \mu(\mu^N - \lambda^N)\frac{\lambda}{\nu}} = \frac{\overline{p}\mu(\mu^N - \lambda^N)}{\mu^{N+1}\left(\overline{p} + \frac{\lambda}{\nu}\right) - \lambda^{N+1}\left(\overline{p} + \frac{\mu}{\nu}\right)} = \frac{\overline{p}}{K_Y}.
\end{aligned}$$

The mean number of customers in the system $\overline{L}_0$ and the mean number of waiting customers $\overline{L}$ are not the same as in the usual M/M/1/N-1-system. They are given by

$$\begin{aligned}
\overline{L}_0 &= \sum_{n=1}^{N}\sum_{k\geq 1} n\pi(n,k) + \sum_{n=1}^{N-1} n\pi(n,0) \\
&= K^{-1}\left(\overline{p} + \frac{\lambda}{\nu}\right)\sum_{n=1}^{N-1} n\left(\frac{\lambda}{\mu}\right)^n + K^{-1}N\overline{p}\left(\frac{\lambda}{\mu}\right)^N \\
&= \frac{\left(\overline{p} + \frac{\lambda}{\nu}\right)\frac{\lambda}{\mu}\left(\mu^N - N\lambda^{N-1}\mu + (N-1)\lambda^N\right)\frac{\mu^2}{(\mu-\lambda)^2} + N\lambda^N\overline{p}}{\mu^N K} \\
&= \frac{\left(\overline{p} + \frac{\lambda}{\nu}\right)\frac{\lambda}{\mu}\left(\mu^N - N\lambda^{N-1}\mu + (N-1)\lambda^N\right)\frac{\mu^2}{\mu-\lambda} + N\lambda^N(\mu-\lambda)\overline{p}}{\mu^N(\mu-\lambda) K} \\
&= \frac{\left(\overline{p} + \frac{\lambda}{\nu}\right)\lambda\mu^{N+1}}{\mu^N(\mu-\lambda)^2K} - \frac{\left(\overline{p} + \frac{\lambda}{\nu}\right)N\lambda^N\mu^2}{\mu^N(\mu-\lambda)^2K} + \frac{\left(\overline{p} + \frac{\lambda}{\nu}\right)(N-1)\lambda^{N+1}\mu}{\mu^N(\mu-\lambda)^2K} + \frac{N\lambda^N(\mu-\lambda)\overline{p}}{\mu^N(\mu-\lambda)K} \\
&= \frac{\lambda(\mu^N - \lambda^N)}{\mu^{N-1}(\mu-\lambda)^2K}\left(\overline{p} + \frac{\lambda}{\nu}\right) - \frac{N\lambda^{N+1}}{\mu^N(\mu-\lambda)K}\left(\overline{p} + \frac{\mu}{\nu}\right),
\end{aligned} \tag{79}$$

$$\begin{aligned}
\overline{L} &= \overline{L}_0 - \mathsf{P}(X > 0) \\
&= \overline{L}_0 - \left(1 - K^{-1}\left(\overline{p} + \frac{\lambda}{\nu}\right)\right) \\
&= \frac{\left(\overline{p} + \frac{\lambda}{\nu}\right)\left(\lambda\mu\frac{\mu^N - \lambda^N}{\mu-\lambda} + \mu^N(\mu-\lambda)\right) - N\lambda^{N+1}\left(\overline{p} + \frac{\mu}{\nu}\right) - \mu^N(\mu-\lambda)K}{\mu^N(\mu-\lambda)K}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\left(\overline{p}+\frac{\lambda}{\nu}\right)\left(\lambda\mu\frac{\mu^N-\lambda^N}{\mu-\lambda}+\mu^N(\mu-\lambda)\right)-N\lambda^{N+1}\left(\overline{p}+\frac{\mu}{\nu}\right)}{\mu^N(\mu-\lambda)K} \\
&\quad -\frac{\mu^{N+1}\left(\overline{p}+\frac{\lambda}{\nu}\right)-\lambda^{N+1}\left(\overline{p}+\frac{\mu}{\nu}\right)}{\mu^N(\mu-\lambda)K} \\
&= \frac{\lambda^2(\mu^{N-1}-\lambda^{N-1})}{\mu^{N-1}(\mu-\lambda)^2K}\left(\overline{p}+\frac{\lambda}{\nu}\right)-\frac{(N-1)\lambda^{N+1}}{\mu^N(\mu-\lambda)K}\left(\overline{p}+\frac{\mu}{\nu}\right). \qquad (80)
\end{aligned}
$$

Using Little's formula the mean sojourn time $\overline{W}_0$ and the mean waiting time $\overline{W}$ of the customers are given by

$$
\begin{aligned}
\overline{W}_0 &= \frac{\overline{L}_0}{\lambda_A}=\frac{\mu^N-\lambda^N}{\mu^{N-1}(\mu-\lambda)^2\overline{p}K_X}\left(\overline{p}+\frac{\lambda}{\nu}\right)-\frac{N\lambda^N}{\mu^N(\mu-\lambda)\overline{p}K_X}\left(\overline{p}+\frac{\mu}{\nu}\right) \\
&= \frac{1}{(\mu-\lambda)\overline{p}}\left(\overline{p}+\frac{\lambda}{\nu}\right)-\frac{N\lambda^N}{\mu(\mu^N-\lambda^N)\overline{p}}\left(\overline{p}+\frac{\mu}{\nu}\right) \\
&= \frac{1}{\mu-\lambda}\left(1+\frac{\lambda}{\nu\overline{p}}\right)-\frac{N\lambda^N}{\mu(\mu^N-\lambda^N)}\left(1+\frac{\mu}{\nu\overline{p}}\right), \\
\overline{W} &= \frac{\overline{L}}{\lambda_A}=\frac{\lambda(\mu^{N-1}-\lambda^{N-1})}{\mu^{N-1}(\mu-\lambda)^2\overline{p}K_X}\left(\overline{p}+\frac{\lambda}{\nu}\right)-\frac{(N-1)\lambda^N}{\mu^N(\mu-\lambda)\overline{p}K_X}\left(\overline{p}+\frac{\mu}{\nu}\right) \\
&= \frac{\lambda(\mu^{N-1}-\lambda^{N-1})}{(\mu-\lambda)(\mu^N-\lambda^N)\overline{p}}\left(\overline{p}+\frac{\lambda}{\nu}\right)-\frac{(N-1)\lambda^N}{\mu(\mu^N-\lambda^N)\overline{p}}\left(\overline{p}+\frac{\mu}{\nu}\right). \qquad \square
\end{aligned}
$$

Proof of Theorem 5.1

$$
\begin{aligned}
s &= \sum_{k\geq 1}k\mathsf{P}^0_{N_R}(Y_{net-}=k)=0, \\
\alpha_1 &= \mathsf{P}^0_{N_R}(\max\{0,Y_{0-}-X_{0-}\}>0)=0. \qquad \square
\end{aligned}
$$

Proof of Theorem 5.2 We carry out the proof for the M/M/1/∞-system with (r,Q)- and (r,S)-policy at the same time. With the second definition of the net inventory position (42) we obtain

$$
\begin{aligned}
s &= \sum_{k=1}^{\infty}k\mathsf{P}^0_{N_R}(Y_{net,0-}=k)=\sum_{k=1}^{\infty}k\mathsf{P}^0_{N_R}(\max\{0,Y_{0-}-X_{0-}\}=k) \\
&= \sum_{h=1}^{\infty}\sum_{k=1}^{h}k\mathsf{P}^0_{N_R}(X_{0-}=h-k,Y_{0-}=h)=\frac{\nu}{\lambda_R}K^{-1}\sum_{h=1}^{r}\sum_{k=1}^{h}k\left(\frac{\lambda}{\mu}\right)^{h-k}C(h).
\end{aligned}
$$

For the (r,Q)- and the (r,S)-system $\frac{\nu}{\lambda_R}K^{-1}=\frac{\nu}{\lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r K_X^{-1}$ with $K_X=\frac{\mu}{\mu-\lambda}$. Furthermore from Theorems 3.3 and 4.3 the constants $C(h)$, $1\leq h\leq r$ are the same. This yields

$$
\begin{aligned}
s &= \frac{\nu}{\lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r K_X^{-1}\sum_{h=1}^{r}C(h)\left(\frac{\lambda}{\mu}\right)^{h-1}\sum_{k=1}^{h}k\left(\frac{\mu}{\lambda}\right)^{k-1} \\
&= \frac{\nu}{\lambda K_X}\left(\frac{\lambda}{\lambda+\nu}\right)^r\sum_{h=1}^{r}C(h)\left(\frac{\lambda}{\mu}\right)^{h-1}\left(1-(h+1)\left(\frac{\mu}{\lambda}\right)^h+h\left(\frac{\mu}{\lambda}\right)^{h+1}\right)\frac{\lambda^2}{(\mu-\lambda)^2}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\nu}{\mu-\lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r \sum_{h=1}^{r} C(h)\left(\left(\frac{\lambda}{\mu}\right)^h - (h+1) + h\frac{\mu}{\lambda}\right) \\
&= \frac{\nu}{\mu-\lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r \left\{\frac{\lambda}{\mu-(\lambda+\nu)}\left(1-\left(\frac{\lambda+\nu}{\mu}\right)^r\right) + \frac{\mu-\lambda}{\lambda}\sum_{h=1}^{r} hC(h)\right. \\
&\quad \left. +\frac{\lambda}{\nu}\left(1-\left(\frac{\lambda+\nu}{\lambda}\right)^r\right)\right\} \\
&= \frac{\nu}{\mu-\lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r \left\{\frac{\lambda}{\mu-(\lambda+\nu)} - \frac{\lambda}{\mu-(\lambda+\nu)}\left(\frac{\lambda+\nu}{\mu}\right)^r\right. \\
&\quad \left. +\frac{\mu-\lambda}{\lambda}\frac{\lambda^2}{\nu^2}\left(1-(r+1)\left(\frac{\lambda+\nu}{\lambda}\right)^r + r\left(\frac{\lambda+\nu}{\lambda}\right)^{r+1}\right) + \frac{\lambda}{\nu} - \frac{\lambda}{\nu}\left(\frac{\lambda+\nu}{\lambda}\right)^r\right\} \\
&= \frac{\lambda\nu}{(\mu-\lambda)(\mu-(\lambda+\nu))}\left(\frac{\lambda}{\lambda+\nu}\right)^r - \frac{\lambda\nu}{(\mu-\lambda)(\mu-(\lambda+\nu))}\left(\frac{\lambda}{\mu}\right)^r \\
&\quad +\frac{\lambda}{\nu}\left(\frac{\lambda}{\lambda+\nu}\right)^r - \frac{\lambda}{\nu}(r+1) + \frac{\lambda}{\nu}r\frac{\lambda+\nu}{\lambda} + \frac{\lambda}{\mu-\lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r - \frac{\lambda}{\mu-\lambda} \\
&= r - \frac{\lambda}{\nu}\left(1-\left(\frac{\lambda}{\lambda+\nu}\right)^r\right) - \frac{\lambda}{\mu-\lambda}\alpha_1,
\end{aligned}
$$

where α_1 is the α_1-service level for the net inventory position as defined in (42). It is derived next. With the same reasoning as above the α_1-service level of the two systems with the second definition of Y_{net} is given by

$$
\begin{aligned}
\alpha_1 &= \mathsf{P}^0_{N_R}(Y_{0-} - X_{0-} > 0) = \sum_{h=1}^{r} \mathsf{P}^0_{N_R}(X_{0-} < h, Y_{0-} = h) \\
&= \frac{\nu}{\lambda_R}K^{-1}\sum_{h=1}^{r}\sum_{n=0}^{h-1}\left(\frac{\lambda}{\mu}\right)^n C(h) \\
&= \frac{\nu}{\lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r \frac{1}{K_X}\sum_{h=1}^{r} C(h)\left(1-\left(\frac{\lambda}{\mu}\right)^h\right) K_X \\
&= -\left(\frac{\lambda}{\lambda+\nu}\right)^r\left(1-\left(\frac{\lambda+\nu}{\lambda}\right)^r\right) - \frac{\nu}{\lambda}\left(\frac{\lambda}{\lambda+\nu}\right)^r \frac{\lambda}{\mu}\sum_{h=1}^{r} C(h)\left(\frac{\lambda}{\mu}\right)^{h-1} \\
&= -\left(\frac{\lambda}{\lambda+\nu}\right)^r\left(1-\left(\frac{\lambda+\nu}{\lambda}\right)^r\right) - \frac{\nu}{\mu}\left(\frac{\lambda}{\lambda+\nu}\right)^r \frac{1-\left(\frac{\lambda+\nu}{\mu}\right)^r}{\mu-(\lambda+\nu)}\mu \\
&= 1-\left(\frac{\lambda}{\lambda+\nu}\right)^r - \frac{\nu}{\mu-(\lambda+\nu)}\left(\frac{\lambda}{\lambda+\nu}\right)^r + \frac{\nu}{\mu-(\lambda+\nu)}\left(\frac{\lambda}{\mu}\right)^r \\
&= 1-\frac{1}{\mu-(\lambda+\nu)}\left((\mu-\lambda)\left(\frac{\lambda}{\lambda+\nu}\right)^r - \nu\left(\frac{\lambda}{\mu}\right)^r\right). \qquad \square
\end{aligned}
$$

Proof of Corollary 5.3. Define a function g as follows

$$
\begin{aligned}
g : \mathbb{N}_0 &\rightarrow \mathbb{R}, \\
r &\mapsto \frac{1}{\mu-(\lambda+\nu)}\left((\mu-\lambda)\left(\frac{\lambda}{\lambda+\nu}\right)^r - \nu\left(\frac{\lambda}{\mu}\right)^r\right).
\end{aligned}
$$

Then, clearly $\alpha_1(r) = 1 - g(r)$. In order to show the assertion it suffices to proof that $\left(\frac{\lambda}{\mu}\right)^r \leq g(r) \leq 1$ and that g is decreasing in r. We proof the first assertion by induction.

Beginning of the induction: $g(0) = 1 = \left(\frac{\lambda}{\mu}\right)^0$.

$$\begin{aligned} g(1) &= \frac{1}{\mu - (\lambda+\nu)} \left((\mu-\lambda)\frac{\lambda}{\lambda+\nu} - \nu\frac{\lambda}{\mu} \right) = \frac{\lambda}{\mu} \frac{(\mu-\lambda)\mu - \nu(\lambda+\nu) - \nu\mu + \nu\mu}{(\mu-(\lambda+\nu))(\lambda+\nu)} \\ &= \frac{\lambda}{\mu}\left(\frac{\mu}{\lambda+\nu} + \frac{\nu}{\lambda+\nu} \right) = \frac{\lambda}{\mu}\frac{\mu+\nu}{\lambda+\nu}. \end{aligned}$$

$g(1) \geq \frac{\lambda}{\mu}$ and $g(1) < 1$ hold because of the ergodicity condition $\lambda < \mu$.
Induction assumption: $\left(\frac{\lambda}{\mu}\right)^{r-1} \leq g(r-1) \leq 1$ for some $r \in \mathbb{N}$.
Induction step:

$$\begin{aligned} g(r) &= \frac{1}{\mu-(\lambda+\nu)} \left((\mu-\lambda)\left(\frac{\lambda}{\lambda+\nu}\right)^r - \nu\left(\frac{\lambda}{\mu}\right)^r \right) \\ &= \frac{1}{\mu-(\lambda+\nu)} \left((\mu-\lambda)\left(\frac{\lambda}{\lambda+\nu}\right)^{r-1} - \nu\left(\frac{\lambda}{\mu}\right)^{r-1} \right) \frac{\lambda}{\lambda+\nu} \\ &\quad + \frac{1}{\mu-(\lambda+\nu)}\nu\left(\frac{\lambda}{\mu}\right)^{r-1}\left(\frac{\lambda}{\lambda+\nu} - \frac{\lambda}{\mu}\right) \\ &= \frac{\lambda}{\lambda+\nu} g(r-1) + \frac{\nu}{\lambda+\nu}\left(\frac{\lambda}{\mu}\right)^r. \end{aligned}$$

$\left(\frac{\lambda}{\mu}\right)^{r-1} \leq g(r-1)$ yields $\left(\frac{\lambda}{\mu}\right)^r \leq g(r)$ since

$$g(r) \geq \frac{\lambda}{\lambda+\nu}\left(\frac{\lambda}{\mu}\right)^{r-1} + \frac{\nu}{\lambda+\nu}\left(\frac{\lambda}{\mu}\right)^r > \frac{\lambda+\nu}{\lambda+\nu}\left(\frac{\lambda}{\mu}\right)^r = \left(\frac{\lambda}{\mu}\right)^r.$$

$g(r) \leq 1$ follows from $g(r-1) \leq 1$ and $\left(\frac{\lambda}{\mu}\right)^r \leq 1$.
Now we show that g is strictly decreasing in r.

$$\begin{aligned} g(r-1) - g(r) &= g(r-1)\left(1 - \frac{\lambda}{\lambda+\nu}\right) - \frac{\nu}{\lambda+\nu}\left(\frac{\lambda}{\mu}\right)^r \\ &= \frac{\nu}{\lambda+\nu}\left(g(r-1) - \left(\frac{\lambda}{\mu}\right)^r \right) \\ &\geq \frac{\nu}{\lambda+\nu}\left(\frac{\lambda}{\mu}\right)^{r-1}\left(1 - \frac{\lambda}{\mu}\right) > 0 \text{ for all } r \in \mathbb{N}. \end{aligned}$$

□

References

[BB03] F. Baccelli and P. Brémaud. *Elements of Queueing Theory*. Springer, New York, 2003.

[BK99] O. Berman and E. Kim. Stochastic models for inventory management at service facilities. *SM*, 15(4):695 – 718, 1999.

[BK01] O. Berman and E. Kim. Dynamic order replenishment policy in internet-based supply chains. *MMOR*, 53:371 – 390, 2001.

[BL85] D.J. Buchanan and R.F. Love. A (Q,R) inventory model with lost sales and Erlang-distributed lead times. *Naval Research Logistics Quaterly*, 32:605 – 611, 1985.

[BS00] O. Berman and K.P. Sapna. Inventory management at service facilities for systems with arbitrarily distributed service times. *SM*, 16(3 & 4):343 – 360, 2000.

[BS01] O. Berman and K.P. Sapna. Optimal control of service for facilities holding inventory. *Computers & Operations Research*, 28:429 – 441, 2001.

[BS02] O. Berman and K.P. Sapna. Optimal service rates of a service facility with perishable inventory items. *Naval Research Logistics*, 49:464 – 482, 2002.

[BvKM92] P. Brémaud, R. van Kannurpatti, and R. Mazumdar. Event and time averages: A review. *Advances in Applied Probability*, 24:377 – 411, 1992.

[CS99] F.M. Cheng and S.P. Sethi. Optimality of state-dependent (s,S) policies in inventory models with Markov-modulated demand and lost sales. *Production and Operations Management*, 8:183 – 192, 1999.

[GH85] D. Gross and C.M. Harris. *Fundamentals of Queueing Theory. Second Edition.* John Wiley & Sons, New York, 1985.

[HC84] A.C. Hax and D. Candea. *Production and Inventory Management.* Prentice Hall, Englewood Cliffs, N.J., 1984.

[HW63] G. Hadley and T.M. Whitin. *Analysis of Inventory Systems.* Prentice Hall, Englewood Cliffs, N.J., 1963.

[MP98] E. Mohebbi and M.J.M. Posner. A continuous-review inventory system with lost sales and variable lead time. *Naval Research Logistics*, 45:259 – 278, 1998.

[Sch81] H. Schneider. Effect of service-levels on order-points or order-levels in inventory models. *International Journal of Production Research*, 19:615 – 631, 1981.

[SM81] K.C. Sevcik and I. Mitrani. The distribution of queueing networks states at input and output instants. *ACM*, 28:358 – 371, 1981.

[SP85] E.A. Silver and R. Peterson. *Decision Systems for Inventory Management and Production Planning.* John Wiley and Sons, Inc., Chichester – New York – Brisbane – Toronto – Singapore, 1985.

[SSL92] K. Sigman and D. Simchi-Levi. Light traffic heuristic for an M/G/1 queue with limited inventory. *Annals of Operations Research*, 40:371 – 380, 1992.

[Sze95] R. Szekli. *Stochastic Ordering and Dependence in Applied Probability.* Springer-Verlag, New York – Berlin – Heidelberg, 1995.

[Tem85] H. Tempelmeier. Inventory control using a service constraint on the expected customer order waiting time. *EJOR*, 19:313 – 323, 1985.

[Tem00] H. Tempelmeier. Inventory service-levels in the customer supply chain. *ORS*, 22:361 – 380, 2000.

[Ter94] R.J. Tersine. *Principles of Inventory and Materials Management.* PTR Prentice Hall, Englewood Cliffs, N.J., 4 edition, 1994.

[Zhe92] Y.-S. Zheng. On properties of stochastic inventory systems. *Management Science*, 38:87 – 103, 1992.