UNIVERSITÄT HAMBURG

Queueing systems with inventory management with random lead times and with backordering

Maike Schwarz, Hans Daduna

Preprint–No. 2005-03 March 2005

FACHBEREICH MATHEMATIK

SCHWERPUNKT MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

Queueing systems with inventory management with random lead times and with backordering

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

March 2005

Abstract

We investigate $M/M/1/\infty$-systems with inventory management, continuous review, exponentially distributed lead times and backordering. We compute performance measures and derive optimality conditions under different order policies. For performance measures, which are not explicitly at hand, we present an approximation scheme for all possible parameter combinations. Although we cannot completely determine analytically the steady state probabilities for the system we are able to derive functional relations between interesting probabilities and show surprising insensitivity properties of several performance measures.
For the approximations we develop an algorithm adapted to the system structure which suggests easy adaption to other systems.

MSC 2000 Subject Classification: 60K25, 90B22, 90B05

Key words: queueing systems; queue lengths; performance analysis; inventory systems; inventory policy; service level; inventory level;

1 Introduction

This paper is devoted to the problem of determining performance measures for service facilities where demand for single items from an attached inventory occurs when a customer is served. It is usually assumed by inventory models that inventory is depleted at the demand rate (if items are in stock) or at an infinite rate (if demand is backordered). This assumption is realistic for production organizations. However, it is less reasonable for organizations where inventory is required to perform a service. Here inventory is depleted at the service rate (which is higher than the demand rate) during the time customers are queued up awaiting service. The problem is: How are the classical performance measures (e.g. queue length, waiting time, etc.) influenced by the management of an attached inventory? – and vice versa: How has inventory management to react to queueing of customers at the the service facility?

Research concerning integrated queueing-inventory models and their explicit analytical solution started only recently. The common approach is: Define a Markovian system process and then use standard optimization methods to find the optimal control strategy of the inventory or at least structural properties of the optimal policies. Some contributions in this direction are: Sigman and Simchi-Levi [SSL92] derived a light traffic heuristic to find performance descriptors for models, in which the interaction of queueing for service and inventory control is incorporated, and in a sequence of papers Berman and his coauthors ([BK99], [BK01], [BK04], [BS00], [BS02]) defined for

*Work supported by Deutscher Akademischer Austauschdienst and KBN, Project D/02/32206

various systems a Markov process description and then used classical optimization methods to find optimal control strategies for the inventory.
The integrated models found in the literature combine single server queueing systems with an attached 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.
$M/M/1/\infty$-servers with attached inventory under lost sales are investigated in Schwarz et. al. [SSD+03b]. Under lost sales regime newly arriving customers are lost when they arrive at a time when the inventory is depleted. It was shown that for different inventory management policies the stationary distribution is of product form for the case of exponentially distributed replenishment lead time. Moreover, it is shown in Schwarz, Sauer and Daduna [SSD03a] that a server with attached inventory can be integrated into standard separable networks if no customers are allowed to join this server's queue while the inventory is empty. For these systems the equilibrium probabilities have been derived analytically and explicit formulas for the equilibrium and asymptotic average costs and revenues have been provided.

The aim of our research in this paper is to provide explicit performance measures for a class of service systems of M/M/1-type with an attached inventory under continuous review and backordering. 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. The lead time is exponentially distributed with parameter ν.

Berman and Kim [BK99] assumed exponentially distributed inter-arrival and service times and zero replenishment lead time. Their result is that for prescribed order size Q, the optimal policy is to place an order when the inventory is empty and the queue length exceeds some threshold value. One can expect that in the case of random lead times an order should already be placed if there are still some items in stock. This is made precise in [BK01] for exponential or Erlangian replenishment lead times. For prescribed order size Q the optimal policy is to place an order when the queue length exceeds some threshold value given the inventory level.

We consider standard policies which are easy to implement. We concentrate on the well established (r, Q) reorder policies which are investigated in a queueing-inventory model in [SSL92] for light traffic. Motivated by the zero-lead-time results of Berman and Kim we incorporate the possibility of a moderate threshold $N = 1$ for the queue length in case of $(0, Q)$ policy. We derive the main performance measures from pure queueing and inventory theory and study their interconnection in the present model. Since some of the performance measures are not at hand explicitly (in contrast to the lost sales case) we develop appropriated approximations.

Our final proposal is to interpolate two different approximation procedures which both are of independent interest: We combine the light traffic heuristic of Sigman and Simchi-Levi [SSL92] (for a subspaces of the parameter space) and a newly developed approximation which is in principle valid for the whole parameter space but which is outperformed by the Sigman and Simchi-Levi approximation on special regions of the parameter space. We demonstrate the precision of the approximations with the aid of a large sample of test systems ranging over the whole admissable parameter space. Admissibility here refers to stability conditions of the systems and (when it is used by the algorithm) to the light traffic region for the Sigman and Simchi-Levi approximation.

The newly developed approximation to a certain extend resembles a wellknown procedure to obtain quantitative values for steady state probabilities, when explicit formulas cannot be obtained analytically: After transforming via generating functions or Laplace-Stieltjes transforms the steady state equations into some transform domain we obtain functional equations which can be solved up to some boundary probabilities. These boundary probabilities are then computed numerically (exact or via some approximation) and used as an anchor for obtaining the complete equilibrium probabilities. A wellknown procedure is *root finding* when the boundary probabilities are given as the solutions of polynomial equations of high order or using *spectral expansions.* A short summary of available methods is presented in the Appendix A.1. We are able to avoid transform methods and evaluate directly functional relations for the relevant probabilities and performance measures. These are of striking simplicity and furthermore reveal surprising structural properties of the queueing-inventory systems. The so obtained functional relations for the performance measures depend on a rather small set of unknown boundary probabilities. Typically these are probabilities located on the zero queue length boundary of the respective state spaces. In principle applying *root finding* or using *spectral expansions* would be possible. But we substitute this by a simple approximation obtained from related small models where steady states are explicitly obtainable. The observation made in comparing the results with simulations is that we need only very small approximating systems.

We concentrate our presentation for full details on the case of $(0, Q)$-policy with threshold $N = 1$ and on its counterpart the $(0, Q)$-policy without threshold. This enables us to compare the respective performance behaviour and to decide under which parameter cofiguration it is better to wait until at least one customer is present at the server before a replenishment order is placed and in which situations an order should be placed even if no customers are waiting in queue.
For (r, Q)-policy with $r \geq 1$ and (r, S)-policy we refer to [Sch04][Section 1.5.6 and 1.5.7].

Although we developed numerical procedures for a dedicated class of models, we believe that a similar approach can be applied to many other complex systems where the steady state probabilities are not explicitly at hand. We expect that an interplay of explicit functional relations and adapted numerical procedures as developed in this paper will be a promising path for further research.

It should be noticed that the functional relations we derive for the performance measures depend on functional relations and invariance properties for steady state probabilities of the systems. These functional relations and invariance properties seem to be rather surprising. Among others we prove that the steady state probability of being busy for the systems under $(0, Q)$ policy with and without threshold is the same as that of a conventional $M/M/1/\infty$ queue, so this probability neither depends on the lead time characteristics nor on the value of Q !

The rest of the paper is organized as follows. We start by introducing the single server system with inventory management in Section 2 and prove stability conditions for all the systems under consideration. Section 3 is devoted to the $(0, Q)$-policy with threshold $N = 1$: We derive here the functional equations and the performance measures of interest and end with numerical recipies. We discuss in detail the goodness of the new approximations compared to other algorithms. Results for the system without threshold and how it compares to the system with threshold can be found in Section 4. In Section 5 we comment on our findings and indicate directions of further research.

Throughout the paper we assume that all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathsf{P})$. Notation: $\mathbb{N} := \{1, 2, 3, \dots\}$, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, $\mathbb{R}$ are the real numbers, $\mathbb{R}_+ := [0, \infty)$. We set empty sums to 0. $\mathbf{A}^T$ is the transpose of matrix $\mathbf{A}$.

2 Single server system with inventory management

Definition 2.1 (The general system) *At a service system with an attached inventory customers of stochastically identical behaviour arrive one by one and require service. There is a single server with infinite waiting room under first come, first served (FCFS) regime. Each customer needs exactly one item in the inventory for service. If the server is ready to serve a customer which is at the head of the line and there is no item of inventory this service starts at the time instant*

(and then immediately) when the next replenishment arrives at the inventory. Customers arriving during a period when the server waits for the replenishment order enter the end of the queue and are backordered.
A served customer departs immediately from the system and the associated item is removed from the inventory at this time instant as well. If there is another customer in the line and at least one further item in the inventory, the next service starts immediately.

There are costs connected with operating the system originating from both, the queueing of the customers and from holding inventory at the system. We have a fixed holding cost h *per item and time unit in the inventory, a fixed ordering cost* A *for each replenishment order, a cost* v_w *per customer and time unit in the waiting room, and a cost* v_l *per customer and time unit in service. Whenever a customer's service is completed, a revenue* R *is payed to the system.*

There is a policy specified which at each decision point determines whether a replenishment order is placed or not, and how many items are ordered. Decision epochs are arrival and departure instants and arrival times of orders. We always assume that there is at most one outstanding order.

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* $t \geq 0$*. We denote the joint queue length and inventory process by* $Z = ((X(t), Y(t)), t \geq 0)$*. The state space of* Z *is* $E_Z = \{(n,k) : n \in \mathbb{N}_0, k \in \{0, ..., M\}\}$*, where* M *is the maximal size of the inventory, which depends on the order policy, see Definition 2.3. We shall henceforth refer to* Z *as the queueing-inventory process.*

E_Z *will be equipped with the lexicographic order* $(n,k) < (m,l) :\Leftrightarrow [n < m \text{ or } (n = m, k < l)]$*, and we partition the state space* $E_Z = \{(n,k) : n \in \mathbb{N}_0, k \in \{0, ..., M\}\}$ *into subsets* $E_n := \{(n,k) : k \in \{0, ..., M\}\}$*. The index* $n = 0, 1, ...$ *is the level.* The second component of the state pair (n,k) is the *inter-level component* or *phase* of the state.

The system defined so far is investigated under different assumptions on the stochastic behaviour of the driving processes and on the cost functions in a series of papers by Berman, Kim, and Sapna [BK99], [BK01], [BS00] and [BS02]. A related model is investigated by Sigman and Simchi-Levi [SSL92]. For the case of lost sales rule for customer arriving during a stock-out period a similar sytem is investigated by Schwarz et. al. [SSD$^+$03b].

Definition 2.2 (Assumption on the random behaviour of the system) *For the service system with inventory management from Definition 2.1 we assume:*
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* $\mu > 0$*.*
The replenishment lead time is exponentially distributed with parameter $\nu > 0$*. Service times, inter-arrival times, and lead times constitute an independent family of random variables.*

For a prescribed fixed order quantity $Q > 0$ in the exponential system of Definition 2.2 with $\nu = \infty$, i.e. zero lead time the central result of Berman and Kim [BK99] says that an optimal policy does not place an order unless the inventory is empty and the number of customers exceeds some threshold value (see also Schneeweiß [Sch81][p. 55]). With nonzero (random) lead times the situation changes: It is now better to reorder in advance before the inventory empties. It is proved under different assumptions in [BK01], [BS00] and [BS02], that optimal policies show a threshold behaviour. The method applied to show this is stochastic dynamic optimization.

Our approach to the system and our aims are different. We compute the steady state of the system for standard simple to implement policies, and then use the equlibrium probabilities to minimize asymptotic costs or maximize the overall profit (= revenue - costs). The simplicity of our results originates from using incomplete (simplified) information in decision making.

Definition 2.3 (Reorder policies) *For the service system with inventory management from Definition 2.1 we consider the following policies where in all cases we assume that on each replenishment an order of size of Q units arrives.*
- (r, Q) *policy:* $r \geq 0$ *is the reorder point, i.e an order is placed immediately when the inventory drops down to* r.
- $\Pi(1)$ *policy: This is a* $(0, Q)$ *policy with an additional threshold* 1 *for the queue length as decision variable. Under* $\Pi(1)$ *an order is triggered when the inventory is empty and there is at least one customer in system. So if the inventory becomes empty and there is no further customer in system then the next order is placed at the arrival of the next customer.*

Berman and Kim [BK01], [BK01], and Sigman and Simchi-Levi [SSL92] investigated general(r, S) policies. In this paper we investigate systems with random lead times and reorder point $r = 0$, similar to Berman and Kim [BK99] in the zero lead time case. We additionally take into consideration a threshold for the queue length as decision variable, as suggested by their results. The optimality proposals of Berman and Kim should have consequences for the case of fast replenishments. (Note that zero lead time is our asymptotics $\nu \to \infty$.) Our restriction to small or no threshold is supported by the numerical evaluations of optimal policies by Berman and Kim [BK01], [BK01], which show that under moderate lead times no threshold with respect to queue length occurs.
We should point out that nevertheless Berman and Kim proved that optimal policies are queue length dependent: So our policies will only yield suboptimal results, and can be considered as heuristic decision making with simple policies.

Under this simplifying assumptions we show that the problem of stabilization of the system can be solved and computing the steady state distribution of the queueing-inventory process is possible. Clearly standard event driven simulation or directly solving the balance equations may be applied. But we prove that using structural properties of the systems can reduce the residual numerics to a minimum. We further show that it is possible to use numerical procedures which are adapted to the class of problems under consideration, which leads to precise (and fast) estimation, resp. computation of the residual unknown probabilities.

We will concentrate in our demonstrations on the most interesting case of policy $\Pi(1)$. The other cases are sketched only and we indicate that they can be solved in a similar manner. Some results on the system from Definition 2.1 with inventory management under $(0, Q)$ are discussed in Section 4. For the case of (r, Q) policy with $r \geq 1$ we refer to [Sch04][Section 1.5.6].

Definition 2.4 (M/M/1/∞) *A service system with inventory management from Definition 2.1 under the stochastic assumptions of Definition 2.2 under one of the policies from Definition 2.3 is called an M/M/1/∞ system with inventory management under the respective policy.*

For all policies we are able to determine the exact conditions for stabilization of the systems.

Proposition 2.5 *The queueing-inventory process Z from Definition 2.1 with stochastic assumptions 2.2 under any of the policies from Definition 2.3 is a continuous-time Markov process which is irreducible on state space* $E_Z = \{0, 1, \ldots, r + Q\}$ *unless* $\Pi(1)$ *is used.*
Under (r, Q) Z *admits a stationary probability vector if and only if*

$$\lambda < \frac{\mu Q \nu (\nu + \mu)^r}{Q\nu(\nu + \mu)^r + \mu^{r+1}} \quad \text{or equivalently} \quad \frac{1}{Q\nu} \cdot \left(\frac{\mu}{\nu + \mu}\right)^r + \frac{1}{\mu} < \frac{1}{\lambda}. \tag{1}$$

Z is irreducible under $\Pi(1)$ *on* $E_Z - (0, Q)$ *and admits a stationary probability vector if and only if*

$$\lambda < \frac{\mu Q \nu}{Q\nu + \mu} \quad \text{or equivalently} \quad \frac{1}{Q\nu} + \frac{1}{\mu} < \frac{1}{\lambda}. \tag{2}$$

Note that the stability condition under $\Pi(1)$ and under $(0, Q)$ are identical. We sketch the proof for $\Pi(1)$, the other cases are along similar lines but more envolved [Sch04].

Proof of Proposition 2.5. We give the details for the queueing-inventory process when $\Pi(1)$ is in force. Then Z has infinitesimal generator matrix $\mathbf{Q} = (q(i,j), i, j \in E_Z)$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{A}_1^{(0)} & \mathbf{A}_0 & & & \\ \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & & \\ & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix} \tag{3}$$

where $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{(Q+1)\times(Q+1)}$ with $\mathbf{A}_1^{(0)} = -\mathbf{A}_0$,

$$\mathbf{A}_0 = \begin{pmatrix} \lambda & & \\ & \ddots & \\ & & \lambda \end{pmatrix}, \mathbf{A}_1 = \begin{pmatrix} -\nu-\lambda & & & \nu \\ & -\mu-\lambda & & \\ & & \ddots & \\ & & & -\mu-\lambda \end{pmatrix}, \mathbf{A}_2 = \begin{pmatrix} 0 & & & \\ \mu & 0 & & \\ & \ddots & \ddots & \\ & & \mu & 0 \end{pmatrix}.$$

All other entries are equal to zero. $\mathbf{Q}$ is bounded so Z is regular and irreducible on $E_Z - (0, Q)$. Therefore Z is a *quasi-birth-and-death process* (QBD process) with inter-level generator matrix $\mathbf{A} := \mathbf{A}_0 + \mathbf{A}_1 + \mathbf{A}_2$. From Neuts [Neu81][Theorem 3.1.1] we have that if the inter-level generator matrix $\mathbf{A}$ is irreducible, then Z is positive recurrent if and only if $\rho := (\alpha\mathbf{A}_0\mathbf{e})(\alpha\mathbf{A}_2\mathbf{e})^{-1} < 1$, where α is the stationary probability vector of $\mathbf{A}$. We see that

$$\mathbf{A} := \mathbf{A}_0 + \mathbf{A}_1 + \mathbf{A}_2 = \begin{pmatrix} -\nu & & & & \nu \\ \mu & -\mu & & & \\ & \mu & \ddots & & \\ & & \ddots & \ddots & \\ & & & \mu & -\mu \end{pmatrix}$$

is the infinitesimal generator of a finite, irreducible, continuous time Markov process if $\mu, \nu > 0$ with stationary distribution $\alpha(0) = \frac{\mu}{\mu+Q\nu}$, and $\alpha(k) = \frac{\nu}{\mu+Q\nu}$ for $k = 1, ..., Q$.

Then from $\alpha\mathbf{A}_0\mathbf{e} = \lambda$ and $\alpha\mathbf{A}_2\mathbf{e} = \frac{\mu Q\nu}{\mu+Q\nu}$, Z is ergodic if and only if $\frac{\alpha\mathbf{A}_0\mathbf{e}}{\alpha\mathbf{A}_2\mathbf{e}} = \frac{\lambda(\mu+Q\nu)}{\mu Q\nu} < 1$ □

We now put the following assumption in force for the rest of the paper.

Assumption: We henceforth call a system that satisfies (2) to be stable, resp. that its queueing-inventory process Z is stable, and assume that Z is stable and in equilibrium.

3 M/M/1/∞ system with inventory management under $\Pi(1)$

In this section we study in detail the steady state behaviour of the M/M/1/∞ service system with inventory according to Definition 2.4 under policy $\Pi(1)$, i.e. with $(0, Q)$ policy for the inventory management and an additional threshold 1 for the queue length, which must be reached before an order is placed. Recall that under $\Pi(1)$ the queueing-inventory process does not reach the state $(0, Q)$ since a replenishment order is not triggered if there are no customers present at the server. However, for notational simplicity, we shall use state space E_Z, because all formulas correctly assign steady-state probability zero to this state.

3.1 The stable M/M/1/∞ system under $\Pi(1)$

We first notice a structural connection of the queueing-inventory system to pure queueing models.

Remark 3.1 *The queue length process of $M/M/1/\infty$-system with inventory management under $\Pi(1)$ with backordering is stochastically indistinguishable from an $M/M/1/\infty$-system in which every Qth customer has a special service time: $exp(\nu) \star exp(\mu)$. The system can be considered as an $M/M^{(Q)}/1$ queue with Q cyclic service times (see Iravani and Posner [IP95]). In a cycle, i.e. a sequence of Q customers, with service times $S_0, S_1, .., S_{Q-1}$, we have in our case*

$$S_k = \begin{cases} S+V, & k=0 \\ S, & k=1,2,...,Q-1, \end{cases}$$

where S is the requested service time and V is the lead time. The expected total number of arrivals per mean service time is $\rho = \lambda(1/\mu + 1/(Q\nu))$. So $\rho < 1$ is the natural condition for ergodicity of the system, which coincides with (2) in Proposition 2.5.

The right side of the first equation in (2) determines the critical input rate λ^* to the system. The dependence of λ^* on the various parameters of the model provides us with a readily accessible means of comparing the effects of inventory size, replenishment lead time and service rate.
Considering some special cases sheds light on the asymptotic behaviour of the system.
The Limit $\nu \to \infty$: Asymptotically the replenishment lead time vanishes, and formally the ergodicity criterion of Proposition 2.5 turns to $\lambda < \mu$. The steady state probability distribution is

$$\pi(n,k) = \frac{1}{Q}\left(\frac{\lambda}{\mu}\right)^n \left(1-\frac{\lambda}{\mu}\right), \quad n \geq 1, \ k=1,...,Q, \quad \pi(0,k) = \frac{1}{Q}\left(1-\frac{\lambda}{\mu}\right), \quad k=0,...,Q-1.$$

Hence, this system behaves as an open $M/M/1/\infty$-FCFS system with equally distributed inventory positions. This special case has been investigated by Berman and Sapna [BS00] with finite and infinite waiting rooms. Their system corresponds to our system without additional threshold.
The Limit $Q \to \infty$: Asymptotically there are always an infinite number of items in stock, and the impact of the replenishment lead time can be neglected. The ergodicity criterion is $\lambda < \mu$ and the steady state probability distribution of the number of customers in the system is identical to that of the classical $M/M/1/\infty$-FCFS system.
The Special Case $Q = 1$: The queueing process behaves as an $M/G/1/\infty$-FCFS system with Coxian service time distribution $\exp(\nu) \star \exp(\mu)$. The ergodicity criterion is $\frac{1}{\lambda} > \frac{1}{\nu} + \frac{1}{\mu}$ [GH85].

3.2 Invariance properties of the stationary distributions

Remark 3.1 and the results of Iravani and Posner [IP95] and Berman and Sapna [BK99] indicate that we cannot expect to find closed form expressions for the steady state distribution of the queueing-inventory process. Fortunately enough we are able to compute explicitly some important aggregate probabilities for Z. These will lead to some suprisingly simple formulas for the performance characteristics of the queueing system and the inventory. Furthermore the resulting performance indices exhibit strong invariance properties with respect to several changes in the system. We denote by $Z = (X, Y)$ a random vector which is distributed according to the equilibrium of the queueing-inventory process of the $M/M/1/\infty$-system with inventory management under $\Pi(1)$.

Theorem 3.2 *For the stable queueing-inventory process Z holds*

$$\mathsf{P}(X > 0, Y = 0) = \frac{\lambda}{Q\nu}, \quad \mathsf{P}(X > 0, Y = k) = \frac{\lambda}{Q\mu} \quad \textit{for } k = 1, ..., Q.$$

Hence, the probability that a customer has to wait because the inventory is depleted is explicitly known. Adding the probabilities given in Theorem 3.2 leads to

Corollary 3.3 *For the stable queueing-inventory process holds*

$$\mathsf{P}(X=0) \quad = \quad 1-\lambda\left(\frac{1}{Q\nu}+\frac{1}{\mu}\right) \quad \textit{and} \quad \mathsf{P}(X>0,Y>0)=\frac{\lambda}{\mu}.$$

Remark 3.4 *The formulas from Theorem 3.2 and Corollary 3.3 are striking. They express surprising invariance properties of the system. The event $\{X>0,Y>0\}=\{$ a customer is served $\}$ has a probability that is independent of the inventory size Q and the rate ν at which replenishment orders arrive. Furthermore, the probability that the server is busy is exactly the probability that a standard $M/M/1/\infty$-system with parameters λ,μ is busy. However, we still have to require that $\frac{1}{\lambda}>\frac{1}{Q\nu}+\frac{1}{\mu}$ holds. Further, for the busy system the inventory level is independent of the lead time distribution, and the probability that a customer waits for service due to empty inventory is independent of the service intensity.*
A direct consequence is that the conditional probabilities for the inventory level given that there are customers present at the server are equal to the stationary probability vector α associated to the inter-level generator matrix $\mathbf{A}$ that has been deduced in the proof of Proposition 2.5.

$$\begin{aligned}
\mathsf{P}(Y=0|X>0) &= \frac{\mu}{Q\nu+\mu}=\left(1+\frac{Q\nu}{\mu}\right)^{-1}, \\
\mathsf{P}(Y=k|X>0) &= \frac{\nu}{Q\nu+\mu}=\left(Q+\frac{\mu}{\nu}\right)^{-1}, \quad k=1,...,Q.
\end{aligned}$$

These conditional probabilities are independent of λ, which is the throughput of the system.

Proof of Theorem 3.2. Denote by π the steady state distribution of Z and by π_n the row vector $\pi_n=(\pi(n,0),...,\pi(n,Q))$ associated to level $n\in\mathbb{N}_0$. π satisfies $\pi\mathbf{e}=1$ and $\pi\mathbf{Q}=0$:

$$-\pi_0\mathbf{A}_0+\pi_1\mathbf{A}_2 = \mathbf{0}, \tag{4}$$

$$\pi_{n-1}\mathbf{A}_0+\pi_n\mathbf{A}_1+\pi_{n+1}\mathbf{A}_2 = \mathbf{0}, \quad n\in\mathbb{N}. \tag{5}$$

Adding these equations we get $\pi_0(\mathbf{A}_0-\mathbf{A}_0)+\sum_{n=1}^{\infty}\pi_n(\mathbf{A}_0+\mathbf{A}_1+\mathbf{A}_2)=\mathbf{0}$. This is equivalent to $\sum_{n=1}^{\infty}\pi_n\mathbf{A}=\mathbf{0}$, where $\mathbf{A}$ denotes the inter-level generator matrix introduced in the proof of Proposition 2.5. Hence, $\sum_{n=1}^{\infty}\pi_n$ is a multiple of the probability distribution α obtained in that proof. This yields

$$\nu\mathsf{P}(X>0,Y=0) = \mu\mathsf{P}(X>0,Y=k), \quad k=1,...,Q. \tag{6}$$

Multiplying the balance equations (4), (5) with column vector $\mathbf{e}=(1,...,1)^T$ of length $Q+1$ yields

$$-\pi_0\begin{pmatrix}\lambda\\ \lambda\\ \vdots\\ \lambda\end{pmatrix}+\pi_1\begin{pmatrix}0\\ \mu\\ \vdots\\ \mu\end{pmatrix}=0,\ \pi_{n-1}\begin{pmatrix}\lambda\\ \lambda\\ \vdots\\ \lambda\end{pmatrix}+\pi_n\begin{pmatrix}-\lambda\\ -(\lambda+\mu)\\ \vdots\\ -(\lambda+\mu)\end{pmatrix}+\pi_{n+1}\begin{pmatrix}0\\ \mu\\ \vdots\\ \mu\end{pmatrix}=0,\ n\in\mathbb{N}.$$

This implies

$$\lambda\mathsf{P}(X=n) = \mu\mathsf{P}(X=n+1,Y>0), \quad n\in\mathbb{N}_0. \tag{7}$$

Adding these equations yields $\lambda \sum_{n=0}^{\infty} \mathsf{P}(X=n) = \mu \sum_{n=1}^{\infty} \mathsf{P}(X=n, Y>0)$, or equivalently

$$\frac{\lambda}{\mu} \quad = \quad \mathsf{P}(X>0, Y>0). \tag{8}$$

Adding the equations of (6) leads to

$$\sum_{k=1}^{Q} \mathsf{P}(X>0, Y=k) = \mathsf{P}(X>0, Y>0) = \frac{Q\nu}{\mu} \mathsf{P}(X>0, Y=0).$$

Using (8) we obtain

$$\mathsf{P}(X>0, Y=0) = \frac{\lambda}{Q\nu} \quad \text{and} \quad \mathsf{P}(X>0, Y=k) = \frac{\lambda}{Q\mu} \quad \text{for } k=1,...,Q. \qquad \square$$

Remark 3.5 *Equations* (6) *and* (7) *are partial balance equations for probability flows in the stable system. (6) relate steady-state probabilities for the inventory process. If the process is in level* 0 *no phase change is possible. Therefore, we get from the above computations only insight into phase changes if the number of customers at the server is positive.*

(7) relate steady-state probabilities for the queue length process, i.e. they quantify movements of customers. The asymmetry of the expressions reflects the property that in an $M/M/1/\infty$-system with inventory no customer can leave the server if there is no inventory on hand.

3.3 Performance measures for the queueing-inventory system

Assume that the stable queueing-inventory process Z is stationary with steady state probability π. We will compute the following operating characteristics for the stationary system:

- $\overline{I}$, the average on-hand inventory position.
- λ_R, the expected reorder rate.
- $\overline{W}$, the average waiting time for a customer from her arrival until she enters the server since there are no more customers waiting in front of her. So stalling of the server due to an ongoing lead time for a replenishment while the customer is in the service place is not included.
- α_2-service level, which is with net inventory $Y(t)$ at time t

  $$\alpha_2 \quad := \quad \mathsf{P}\,(\text{net inventory position at the end of a unit of time } > 0).$$

- β-service level,

  $$\beta \quad := \quad \frac{\mathsf{E}(\text{ immediately satisfied demand per cycle })}{\mathsf{E}(\text{ total demand per cycle })}, \tag{9}$$

  'Immediately satisfied demand' occurs when a customer who enters the server can start his service immediately without having to wait for a replenishment order to arrive.

It is important that coupling of queueing and inventory management leads to new aspects when defining performance parameters. To encompass their interplay we have to redefine some of the classical system descriptors. E.g., in pure inventory systems the net inventory position is usually defined as the on-hand inventory minus backorders. Therefore, we might have used $Y_{net} = \max\{0, Y-X\}$ as the net inventory. However, in the context of queueing not all waiting customers are backorders and customers do not expect to be served immediately.

Theorem 3.6 (Performance measures for the queueing-inventory system) *For the stable $M/M/\ 1/\infty$-system with attached inventory, $(0,Q)$-policy, late replenishment and backordering according to Definition 2.1 the measures of system performance are*

$$\begin{aligned}
\overline{I} &= \frac{(Q+1)\lambda}{2\mu} + \sum_{k=1}^{Q-1} k\pi(0,k), && (10)\\
\lambda_R &= \frac{\lambda}{Q}, && (11)\\
\alpha_2 &= 1 - \frac{\lambda}{Q\nu} - \pi(0,0), && (12)\\
\beta &= 1 - \frac{1}{Q}, && (13)\\
\overline{W} &= \frac{\mu}{Q\nu(\mu-\lambda)-\lambda\mu}\left(\frac{Q-1}{2} + \frac{\lambda(Q\nu+\mu)^2}{Q\nu\mu^2} - \overline{I}\right). && (14)
\end{aligned}$$

Proof. The proof of Theorem 3.6 is given in Appendix A.4 on page 30. □

In $\overline{I}$ and $\overline{W}$ the only unknowns are $\pi(0,1),\ldots,\pi(0,Q-1)$, whereas in α_2 we have to find $\pi(0,0)$. Table 1 depicts some limiting values for later comparisons. The first and fourth limits result from

Table 1: Limiting values for performance measures

	$\lim_{\lambda \nearrow \frac{Q\nu\mu}{Q\nu+\mu}}$	$\lim_{\nu\to\infty}$	$\lim_{Q\to\infty}$	$\lim_{Q \searrow \frac{\lambda\mu}{\nu(\mu-\lambda)}}$	$\lim_{Q\to 1}\left(\text{if } \frac{1}{\lambda} > \frac{1}{\mu}+\frac{1}{\nu}\right)$
$\overline{I}$	$\frac{Q(Q+1)\nu}{2(Q\nu+\mu)}$	$\frac{Q-1}{2}+\frac{\lambda}{\mu}$	∞	$\frac{(\lambda\mu+\nu(\mu-\lambda))\lambda}{2\mu\nu(\mu-\lambda)}$	$\frac{\lambda}{\mu}$
λ_R	$\frac{Q\nu+\mu}{\nu\mu}$	$\frac{\lambda}{Q}$	0	$\frac{\nu(\mu-\lambda)}{\mu}$	λ
$\overline{W}$	∞	$\frac{\lambda}{\mu(\mu-\lambda)}$	$\frac{\lambda}{\mu(\mu-\lambda)}$	∞	$\frac{\lambda(\nu^2+\nu\mu+\mu^2)}{\nu\mu(\nu\mu-\lambda(\nu+\mu))}$

$\lim_{\lambda \nearrow \frac{Q\nu\mu}{Q\nu+\mu}} \mathsf{P}(X=0) = \lim_{Q \searrow \frac{\lambda\mu}{\nu(\mu-\lambda)}} \mathsf{P}(X=0) = 0$. The second and third limits are calculated using the steady state probability distributions for $\nu\to\infty$ and $Q\to\infty$ given on p. 7. The second and third limit of $\overline{W}$ are equal to the waiting time in the standard $M/M/1/\infty$-FCFS system. These quantities are lower boundaries of the expected waiting time in the queueing system with inventory management. The last values for $\overline{I}$ and λ_R are obvious. For $\overline{W}$ the limiting value for $Q\to 1$ is just the average waiting time in an $M/G/1/\infty$-FCFS system where the service time has a Coxian distribution given by $\exp(\nu)\star\exp(\mu)$.

Remark 3.7 *We define the total cost function similar to Berman and Kim [BK99], [BK01]*

$$TC = \overline{I}\cdot h + \lambda_R\cdot A + \overline{W}\cdot v_w. \qquad (15)$$

An argument from ergodic theory yields that (15) *is the asymptotic average cost for the system. For asymptotically vanishing replenishment lead time ($\nu\to\infty$) TC is a concave function in Q*

$$TC(Q) = \left(\frac{Q-1}{2}+\frac{\lambda}{\mu}\right)\cdot h + \frac{\lambda}{Q}\cdot A + \frac{\lambda}{\mu(\mu-\lambda)}\cdot v_w.$$

The optimal order quantity is the classical economic order quantity $Q^* = \sqrt{\frac{2A\lambda}{h}}$.
The total profit for the system similar to Berman and Kim [BK04] is

$$TP = \lambda \cdot R - TC = \lambda \cdot R - \overline{I} \cdot h - \lambda_R \cdot A - \overline{W} \cdot v_w. \tag{16}$$

The results in Theorem 3.6 suggest to introduce the variable

$$\overline{L}_0 = \sum_{k=1}^{Q-1} k\, \pi(0,k). \tag{17}$$

$\overline{I}$ and $\overline{W}$ can now be expressed using $\overline{L}_0$ as

$$\overline{I} = \frac{(Q+1)\lambda}{2\mu} + \overline{L}_0, \tag{18}$$

$$\overline{W} = \frac{\mu}{Q\nu(\mu-\lambda) - \lambda\mu} \left(\frac{Q-1}{2} + \frac{\lambda(Q\nu+\mu)^2}{Q\nu\mu^2} - \frac{(Q+1)\lambda}{2\mu} - \overline{L}_0 \right). \tag{19}$$

For later comparisons we state limiting values for $\overline{L}_0$ in Table 2.

Table 2: Limiting values for $\overline{L}_0$

	$\lim\limits_{\lambda \nearrow \frac{Q\nu\mu}{Q\nu+\mu}}$	$\lim\limits_{\nu \to \infty}$	$\lim\limits_{Q \to \infty}$	$\lim\limits_{Q \searrow \frac{\lambda\mu}{\nu(\mu-\lambda)}}$	$\lim\limits_{Q \to 1} \left(\text{if } \frac{1}{\lambda} > \frac{1}{\mu} + \frac{1}{\nu} \right)$
$\overline{L}_0$	0	$\frac{(Q-1)(\mu-\lambda)}{2\mu}$	∞	0	0

3.4 Numerical calculation of the boundary probabilities and approximations

Last step to derive performance measures in Theorem 3.6 is to calculate $\pi_0 = (\pi(0,1), ..., \pi(0,Q-1))$, which can be done by solving the complete steady state equations numerically. We briefly review methods to numerically calculate stationary distributions of the system in the Appendix A.1.

In this section we present two methods to obtain approximations of the performance measures, especially $\overline{I}$ and $\overline{W}$ for the $M/M/1/\infty$-system with backordering and $(0,Q)$ inventory management with threshold 1, i.e. $\Pi(1)$ policy. Note that from the results in Section 3.3 we only have to find an approximation of the quantity $\overline{L}_0$ from (17) in order to approximate $\overline{I}$ and $\overline{W}$. α_1 involves the unknown probability $\pi(0,0)$, an approximation of which is obtained in course of computing $\overline{L}_0$. All other performance measures are known explicitly.
In Section 3.4.1 we use a related closed queueing system with inventory management to find a rule how to distribute the total mass $\mathsf{P}(X=0) = 1 - \lambda\left(\frac{1}{Q\nu} + \frac{1}{\mu}\right)$ from Corollary 3.3 among the inventory positions $0, 1, ..., Q-1$. In Section 3.4.2 we present the light traffic heuristic for the average delay in queue proposed by Sigman and Simchi-Levi [SSL92]. For each approximation we derive limiting values which allow for a first assessment of the quality of the respective approximation. Computational results for the different approximations are given in Section 3.4.3.

3.4.1 Approximation using systems with a limited number of waiting places

In this section the unknown probabilities $\pi(0,k)$ will be approximated using the truncated balance equations for $(\pi^{(N)}(n,k) : n \in \{0,..,N\}, k \in \{0,...,Q\})$ which is the steady state distribution in a corresponding system with a limited number of $N-1$ waiting places or a suitably defined closed network. The solution is used to distribute the correct total mass $1-\lambda\left(\frac{1}{\mu}+\frac{1}{Q\nu}\right)$ found in Corollary 3.3 among the inventory positions $0,...,Q-1$. $\overline{L}_0 = \sum_{k=1}^{Q-1} k\,\pi(0,k)$ is thus approximated by

$$\left(1-\lambda\left(\frac{1}{\mu}+\frac{1}{Q\nu}\right)\right)\left(\sum_{k=0}^{Q-1}\pi^{(N)}(0,k)\right)^{-1}\sum_{k=1}^{Q-1}k\,\pi^{(N)}(0,k). \tag{20}$$

A strong advantage of this method is that we do not need to use the normalized probabilities $\pi^{(N)}(0,k)$ because an additional factor cancels out in (20).

We compute now the nonnormalized solutions of the steady state equations for $N=1,2$ and 3 and the approximations $\overline{L}_0^{CN}$ for $\overline{L}_0$. $C1, C2$ and $C3$ refer henceforth to the number of waiting+service places of the respective approximating systems.

The case $N=1$: The approximation of $\overline{L}_0$ according to (20) is obtained from

$$\pi^{(1)}(0,k) = 1, \quad k=0,...,Q-1, \quad \text{as} \quad \overline{L}_0^{C1} = \frac{(Q-1)(Q\nu\mu-\lambda(Q\nu+\mu))}{2Q\nu\mu}.$$

The case $N=2$: We have

$$\pi^{(2)}(0,k) = 1, \qquad k=0,...,Q-2, \quad \text{and} \quad \pi^{(2)}(0,Q-1) = \frac{\nu}{\lambda+\nu}.$$

Hence, according to (20) the approximation of $\overline{L}_0$ is

$$\overline{L}_0^{C2} = \frac{(Q-1)(Q\nu\mu-\lambda(Q\nu+\mu))}{Q\nu\mu((Q-1)(\lambda+\nu)+\nu)}\left(\frac{(Q-2)(\lambda+\nu)}{2}+\nu\right). \tag{21}$$

The case $N=3$: With $a=\frac{(\lambda+\mu)^2}{\lambda\mu}$, $b=\frac{(\lambda+\mu)(\lambda+\nu)}{\lambda\mu}$, $c=\left(\frac{\lambda+\nu}{\lambda+\mu}+\frac{(\lambda+\mu+\nu)(\lambda^2+\lambda\mu+\mu^2)}{\mu((\lambda+\mu)^2-\mu(\lambda+\nu)a^Q+\mu\nu)}\right)$ we have

$$\pi^{(3)}(0,0)=\frac{\mu}{\lambda}c, \quad \pi^{(3)}(0,k) = \frac{\mu}{\lambda}\left(\frac{1-a^{k+1}}{1-a}c-b\frac{1-a^k}{1-a}\right), \quad k=1,...,Q-2,$$
$$\pi^{(3)}(0,Q-1) = \frac{\nu\mu}{\lambda(\lambda+\mu)},$$

After some calculations the approximation of $\overline{L}_0$ is found to be

$$\overline{L}_0^{C3} = \left(1-\lambda\left(\frac{1}{\mu}+\frac{1}{Q\nu}\right)\right)\frac{\frac{(Q-1)\nu}{\lambda+\mu}+\frac{(Q-2)(Q-1)}{2(1-a)}(c-b)-\frac{(Q-2)a^{Q-1}-(Q-1)a^{Q-2}+1}{(1-a)^3}a(ac-b)}{c+\frac{\nu}{\lambda+\mu}+\frac{Q-2}{1-a}(c-b)-\frac{a(1-a^{Q-2})}{(1-a)^2}(ac-b)}. \tag{22}$$

The quality of these approximations will be evaluated by numerical comparisons and simulations. Their usefulness is supported by the observation that in any case the approximation $\overline{L}_0^{CN}, N=1,2,3$ obtained for $\overline{L}_0$ behaves asymptotically as $\overline{L}_0$, see Table 2. Using (18) and (19) we obtain approximations of the average on-hand inventory $\overline{I}$ and the average waiting time $\overline{W}$ of the customers denoted by $\overline{I}^{CN}$ and $\overline{W}^{CN}$, $N=1,2,3$. More details can be found in Schwarz [Sch04].

3.4.2 Sigman and Simchi-Levi's approximation of the mean waiting time

Sigman and Simchi-Levi [SSL92] proposed a light traffic heuristic for the average delay $\overline{W}^S$ of customers in the system from Definition 2.1 under (r, Q) from Definition 2.3. We want to compare their approximation scheme under the exponential assumptions of Definition 2.2 with the methods developed before. And we will eventually integrate both methods into an interpolation approximation which integrates the advantages of both schemes at different subspaces of the parameter space. At a first glance problems will occur from two points of system specifications:
• Under $\Pi(1)$ we wait until the queue is not empty before we place an order after the inventory emptied, but under $(0, Q)$ we immediately place the order after the inventory emptied.
• $\overline{W}$, the average waiting time for a customer in our system under $\Pi(1)$ is the time from her arrival until she enters the server since there are no more customers waiting in front of her. So eventual stalling of the server due to an ongoing lead time for a replenishment order while the customer is in the service place is not included, while $\overline{W}^S$, the stationary average delay in Sigman and Simchi-Levi's system includes these waiting because of stalling of the server in case of stock-out.

But, simulations revealed that there are strong connections, the proof of the following theorem makes this observation precise: The two differences discussed above cancel out. We consider stationary virtual mean waiting time of customers. The proof is postponed to Appendix A.2.

Theorem 3.8 *The stationary mean waiting time $\overline{W}$ of the ergodic $M/M/1/\infty$-system with inventory management under policy $\Pi(1)$, and the stationary average delay $\overline{W}^S$ under ergodicity in Sigman and Simchi-Levi's system with $r = 0$ and exponential service times with rate μ are the same.*

The heuristics of Sigman and Simchi-Levi [SSL92] is proposed for light traffic which means that the rates satisfy $\nu \leq \lambda \leq \mu$. Applying this approximation to our (exponential) system yields

$$\overline{W}^S = \frac{1}{Q\nu(\mu-\lambda)}\left(\frac{\mu\lambda}{\nu} + \mu + \frac{\lambda\nu}{\mu} + \frac{\lambda(\mu+\nu)}{\mu(\mu-\lambda)}\right).$$

Inserting the approximate $\overline{W}^S$ into (19) and (18) we get further approximations

$$\begin{aligned} \overline{I}^S &= \left(\frac{1}{\mu} - \frac{\lambda}{Q\nu(\mu-\lambda)}\right)\left(\frac{\mu\lambda}{\nu} + \mu + \frac{\lambda\nu}{\mu} + \frac{\lambda(\mu+\nu)}{\mu(\mu-\lambda)}\right) + \frac{Q-1}{2} + \frac{\lambda(Q\nu+\mu)^2}{Q\nu\mu^2}, \\ \overline{L}_0^S &= \overline{I}^S - \frac{(Q+1)\lambda}{2\mu}. \end{aligned} \tag{23}$$

Limiting values of these approximations can be found in Table 3.

Table 3: Limiting values for Sigman and Simchi-Levi's approximation

	$\lim_{Q\to\infty}$	$\lim_{Q\searrow\frac{\lambda\mu}{\nu(\mu-\lambda)}}$	$\lim_{Q\to 1}$
$\overline{I}^S$	∞	$\frac{\lambda\mu}{2\nu(\mu-\lambda)} + \frac{1}{2} + \frac{\lambda}{\mu}$	$\left(\frac{1}{\mu} - \frac{\lambda}{\nu(\mu-\lambda)}\right)\left(\frac{\mu\lambda}{\nu} + \mu + \frac{\lambda\nu}{\mu} + \frac{\lambda(\mu+\nu)}{\mu(\mu-\lambda)}\right) + \frac{\lambda(\nu+\mu)^2}{\nu\mu^2}$
$\overline{W}^S$	0	$\frac{1}{\nu} + \frac{1}{\lambda} + \frac{\nu}{\mu^2} + \frac{\mu+\nu}{\mu^2(\mu-\lambda)}$	$\frac{1}{\nu(\mu-\lambda)}\left(\frac{\mu\lambda}{\nu} + \mu + \frac{\lambda\nu}{\mu} + \frac{\lambda(\mu+\nu)}{\mu(\mu-\lambda)}\right)$
$\overline{L}_0^S$	∞	$\frac{\lambda\mu+\nu(\lambda+\mu)}{2\nu\mu}$	$\left(\frac{1}{\mu} - \frac{\lambda}{\nu(\mu-\lambda)}\right)\left(\frac{\mu\lambda}{\nu} + \mu + \frac{\lambda\nu}{\mu} + \frac{\lambda(\mu+\nu)}{\mu(\mu-\lambda)}\right) + \frac{\lambda(\nu^2+\nu\mu+\mu^2)}{\nu\mu^2}$

The limit $Q \to 1$ can only be considered if $\frac{1}{\lambda} > \frac{1}{\nu} + \frac{1}{\mu}$. Comparison of these results with limiting values for $\overline{L}_0, \overline{I}$ and $\overline{W}$ in Tables 2 and 1 shows that the limits for $Q \searrow \frac{\lambda\mu}{\nu(\mu-\lambda)}$ and $Q \to 1$ do not coincide. If Q is small however, Sigman and Simchi-Levi's approximation gets poor even if λ and ν are very small. For details see Appendix A.3.

3.4.3 Numerical results

In this section, we compare the performance of the different approximation methods presented in Sections 3.4.1 and 3.4.2 for a large number of parameter combinations.

The Test Examples We investigate 288 test examples that are grouped into 4 groups by different arrival rates $\lambda = 0.1, 1/3, 2/3, 0.8$. Each set is composed of 6 subsets by different lead time rates $\nu = 0.01, 0.1, 0.5, 1, 5, 10$. Each subset has 12 different inventory sizes Q = 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50. For all examples $\mu = 1$. Since not all parameter combinations satisfy the ergodicity criterion in (2) the total number of examples reduces to 211.

Exact values for $\overline{L}_0, \overline{I}$ and $\overline{W}$ are presented in Tables 4 – 6 on p. 17 – 19. They have been calculated using the matrix geometric method described in Section A.1. Empty cells correspond to a parameter set which does not satisfy the ergodicity criterion in (2). Several observations can be made based on these results. $\overline{L}_0$ and $\overline{I}$ are both increasing in Q and ν. $\overline{L}_0$ is decreasing in λ, whereas $\overline{I}$ does not imitate this behaviour. $\overline{W}$ is decreasing in Q and ν, and with a lower boundary $\frac{\lambda}{\mu(\mu-\lambda)}$. As is already suggested by this boundary, $\overline{W}$ is strictly increasing in λ.

The relative deviations from the exact value of $\overline{L}_0$ for the different approximation methods are presented in Tables 7 on p. 20 for $C1, C2$ and $C3$. All numerical results for Sigman and Simchi-Levi's approximation are summarized in Table 10 on p. 23.

1. The rel. deviation of $C1$ improves for increasing ν and Q and decreasing λ.

2. The rel. deviations of $C2$ are not monotone in Q. The performance of $C2$ improves rapidly for increasing ν. If $\nu = 5$ all rel. deviations are smaller than 1%. This can again be explained by the more and more equally distributed inventory positions for level 0. Since $C2$ gives the same probability to the states $(0,0), ..., (0, Q-2)$ and a smaller probability to $(0, Q-1)$. As before the rel. deviations get worse if λ increases. $C2$ outperforms $C1$ for all parameter sets.

3. The rel. deviations of $C3$ decrease rapidly for increasing ν and decreasing λ. In general the performance of $C3$ is very good. However, if ν is small the deviations can become quite large (see $\lambda = 0.1$ with $\nu = 0.01$ and $\lambda = 1/3$, $\lambda = 2/3$ and $\lambda = 0.8$ all with $\nu = 0.1$). Apart from three parameter sets $C3$ gives better results than $C2$. If Q is large $C2$ and $C3$ perform equally well. Thus, when compared with $C1$ and $C2, C3$ in general is the best approximation. Since improvements are not large when going from $C2$ to $C3$ we do not derive $C4$.

4. Sigman and Simchi-Levi's approximation is defined only for parameter sets which satisfy $\nu \le \lambda \le \mu$. It outperforms $C1, C2$ and $C3$ if λ and ν are small and Q is large. In total Sigman and Simchi-Levi's approximation constitutes the best approximation in seven cases ($\lambda = 0.1, \nu = 0.01, Q = 30, 40$ or 50 and $\lambda = 1/3, \nu = 0.1, Q = 20, 30, 40$ or 50). If Q is small however, the performance of Sigman and Simchi-Levi's approximation gets poor even if λ and ν are very small. This result has already been anticipated at the end of Section 3.4.2.

The above observations clearly indicate that $C1$, $C2$ and $C3$ improve for increasing values of ν and Q and decreasing values of λ. This results from the right behaviour of these approximations in the limit $\nu \to \infty$ and $Q \to \infty$ (see Table 2). Moreover, the larger ν and Q become the less the impact of replenishment orders on the behaviour of the system and the more equally inventory positions are distributed. Out of these approximation methods, $C3$ performs best. If ν and λ are small and Q is large Sigman and Simchi-Levi's approximation performs better than C3. However, if Q is small Sigman and Simchi-Levi's approximation should not be used.

In a second step we will now investigate approximation $\overline{L}_0^A$ impacts the corresponding approximations $\overline{I}^A$ obtained from (18). Approximations for $\overline{W}^A$ obtained similarly from (19) behave almost parallel and can be found with details in [Sch04].

Fortunately, the decreasing performance of $\overline{L}_0^A, A \in \{C1, C2, C3, S\}$ for larger values of λ is eliminated in approximations of $\overline{I}$ since the constant part $\frac{(Q+1)\lambda}{2\mu}$ is of more and more importance (compare the results in Tables 7 and 8 for $\lambda = 2/3$ and 0.8). If on the other hand λ is small the good performance of approximations of $\overline{I}$ is assured as well. The rel. deviation for an approximation $\overline{I}^A$ of $\overline{I}$ can be expressed as follows by $\overline{L}_0$ and its corresponding approximation $\overline{L}_0^A$

$$\frac{\overline{I} - \overline{I}^A}{\overline{I}} = \frac{\overline{L}_0 - \overline{L}_0^A}{\overline{I}}. \quad \text{Hence, } \frac{\text{rel. deviation}(\overline{I})}{\text{rel. deviation}(\overline{L}_0)} = \frac{\overline{L}_0}{\overline{I}} = \frac{\overline{L}_0}{\frac{(Q+1)\lambda}{2\mu} + \overline{L}_0} < 1.$$

This shows that the rel. deviation of an approximation for $\overline{I}$ is always smaller than that of the corresponding approximation for $\overline{L}_0$. Only for $(\lambda = 0.1, \nu = 0.01, Q = 20)$ the maximal rel. deviation of all approximations is larger than 5 %. In four cases the maximal rel. deviation of all approximations if larger than 3 % ($\lambda = 0.1, \nu = 0.01, Q = 40$ or 50 and $\lambda = 2/3, \nu = 0.1, Q = 40$ or 50).

Although we observed very good performance of $C3$ for large number of parameter sets, $C3$ can have very poor performance if ν is small. This is the region where Sigman and Simchi-Levi's approximation applies. However, Sigman and Simchi-Levi's approximation yields poor results for small Q and large λ. This suggests to integrate both methods into a new approximation.

3.5 Interpolation

The numerical results in Section 3.4.3 show that $C3$ is a very good approximation, which however performs poorly for small values of ν. On the other hand, S performs very well if ν is small but gets worse for increasing values of ν and λ and decreasing values of Q.

This leads to the following interpolation approximation: for parameter values Q, λ, and μ we interpolate $\overline{L}_0$ for values of ν in the interval $\left(\frac{\lambda\mu}{Q(\mu-\lambda)}, \lambda\right]$ where the left boundary is from ergodicity criterion (2) and the right from Sigman and Simchi-Levi's assumptions. Within the interval we use a linear convex combination of Sigman and Simchi-Levi's approximation $\overline{L}_0^S$ and $\overline{L}_0^{C3}$. Outside the interval $C3$ is used to approximate $\overline{L}_0$. However, we make one more adjustment: $\overline{L}_0^S$ tends to overestimate $\overline{L}_0$ for small values of Q. Our numerical results indicate that there is good reason to believe that $\overline{L}_0^{C3}$ is an upper bound of $\overline{L}_0$. Therefore, whenever Sigman and Simchi-Levi's approximation is larger than $\overline{L}_0^{C3}$ we use $\overline{L}_0^{C3}$. This leads to the following

Algorithm: Interpolation Approximation

Given: ν, μ, λ, Q; satisfying (2)
calculate $\overline{L}_0^{C3}$;
If $\nu < \lambda$ then

    calculate $\overline{L}_0^{S}$;
    If $\overline{L}_0^{S} < \overline{L}_0^{C3}$ then

$$g = \frac{Q\nu\mu - \lambda(Q\nu + \mu)}{\lambda((Q-1)\mu - Q\lambda)};$$

$$\overline{L}_0^{I} = (1-g)\cdot \overline{L}_0^{S} + g \cdot \overline{L}_0^{C3};$$

    else

$$\overline{L}_0^{I} = \overline{L}_0^{C3};$$

    endif;

else

$$\overline{L}_0^{I} = \overline{L}_0^{C3};$$

endif.

Numerical results Table 11 on p. 24 depicts the rel. deviation from interpolation approximation $\overline{L}_0^{I}$ to the exact value $\overline{L}_0$ for the parameter settings from Section 3.4.3. Observe from the results that for $\lambda = 0.1$ and $\nu = 0.01$ approximation $C3$ could not be improved for $Q = 20$. For $Q = 30$ $\overline{L}_0^{I}$ is just a little bit worse then the best approximation $\overline{L}_0^{S}$. If $Q = 40$ or 50 however, $\overline{L}_0^{I}$ performs better than any other approximation before. In all other cases $\overline{L}_0^{I} = \overline{L}_0^{C3}$. For $\lambda = 1/3$ $\overline{L}_0^{I}$ is different from $\overline{L}_0^{C3}$ if $\nu = 0.1$ and $Q \geq 20$. For $Q = 20$ $\overline{L}_0^{I}$ outperforms all other approximations. For $Q = 30, 40$ and 50 the performance of $\overline{L}_0^{I}$ is slightly worse than those of $\overline{L}_0^{S}$. If $\lambda = 2/3$ or 0.8 the performance of $\overline{L}_0^{S}$ is poor resulting in $\overline{L}_0^{I} = \overline{L}_0^{C3}$ for all parameter settings. Consequently, the rel. deviations of $\overline{L}_0^{I}$ still exceed 20% for $\nu = 0.1$.

As has been shown in Section 3.4.3 the rel. deviations of $\overline{I}^{I}$ from $\overline{I}$ are smaller than those of $\overline{L}_0^{I}$ from $\overline{L}_0$. Table 12 shows that for $\lambda = 0.1$ $\overline{I}^{I}$ only performs badly if $\nu = 0.01$ and $Q = 20$. For $\lambda = 1/3$ the rel. deviation of $\overline{I}^{I}$ from $\overline{I}$ is maximal at $\nu = 0.1$ and $Q = 50$ with a deviation of only 3%. If $\lambda = 2/3$ or 0.8 the performance of $\overline{I}^{I}$ is very convincing with rel. deviations of more than 3% occuring only twice at $\lambda = 2/3, \nu = 0.1, Q = 40$ and 50.
We conclude that $\overline{I}^{I}$ is a very good approximation of $\overline{I}$. If $\nu \geq 1$ rel. deviations are less than 0.9 %, if $\nu \geq 0.5$ they are less than 1.7% and if $\nu \geq 0.1$ they are still less than 3.9%. However, a high rel. deviation occurs if λ, ν and Q are small simultaneously.

4 $M/M/1/\infty$ system with $(0, Q)$ inventory management policy

The stability condition for the $M/M/1/\infty$-system under $(0, Q)$ inventory management without threshold for the queue length was computed in Proposition 2.5. In this section we compare its performance with that of the $M/M/1/\infty$-system with inventory management under $\Pi(1)$. In order to differentiate between the two systems we add indices l (for LATE ORDERING) under $\Pi(1)$ and e for (EARLY ORDERING) under $(0, Q)$ to the steady state probability distribution as well as to the performance measures for the systems with the respective replenishment policies. Note that state $(0, Q)$ now has a positive steady state probability if Z_e is ergodic. We present analytically obtain-

	Q	$\nu = 0.01$	$\nu = 0.1$	$\nu = 0.5$	$\nu = 1$	$\nu = 5$	$\nu = 10$
$\lambda = 0.1$	3		0.4599	0.7859	0.8402	0.8875	0.8937
	4		0.8163	1.2162	1.2806	1.3357	1.3428
	5		1.1984	1.6535	1.7246	1.7846	1.7923
	6		1.5962	2.0948	2.1706	2.2338	2.2419
	7		2.0048	2.5384	2.6176	2.6833	2.6916
	8		2.4213	2.9836	3.0654	3.1329	3.1414
	9		2.8438	3.4299	3.5137	3.5825	3.5913
	10		3.271	3.8769	3.9623	4.0323	4.0411
	20	2.8089	7.6617	8.3635	8.4562	8.5311	8.5406
	30	6.4404	12.1245	12.859	12.9541	13.0308	13.0404
	40	10.3291	16.6059	17.3567	17.4531	17.5306	17.5403
	50	14.3788	21.0947	21.8554	21.9525	22.0305	22.0402
$\lambda = 1/3$	3			0.3824	0.5094	0.6317	0.6489
	4			0.6590	0.8131	0.9590	0.9793
	5			0.9513	1.1256	1.2883	1.3106
	6		0.2042	1.2527	1.4432	1.6187	1.6425
	7		0.4284	1.5603	1.7641	1.9498	1.9747
	8		0.6668	1.8723	2.0874	2.2813	2.3071
	9		0.9158	2.1876	2.4125	2.6132	2.6398
	10		1.1732	2.5056	2.7388	2.9453	2.9725
	20		4.0007	5.7580	6.0374	6.2728	6.3030
	30		7.0580	9.0611	9.3583	9.6041	9.6353
	40		10.2238	12.3792	12.6854	12.9364	12.9682
	50		13.4489	15.7033	16.0150	16.2691	16.3012
$\lambda = 2/3$	3				0.0962	0.2786	0.3053
	4				0.2204	0.4362	0.4675
	5			0.1104	0.3572	0.5968	0.6310
	6			0.2333	0.5009	0.7588	0.7954
	7			0.3640	0.6488	0.9218	0.9603
	8			0.5001	0.7997	1.0854	1.1254
	9			0.6401	0.9527	1.2494	1.2908
	10			0.7830	1.1074	1.4138	1.4563
	20			2.2941	2.6982	3.0667	3.1163
	30		1.2540	3.8703	4.3232	4.7261	4.7794
	40		2.6004	5.4749	5.9622	6.3881	6.4437
	50		4.0032	7.0958	7.6090	8.0514	8.1088
$\lambda = 0.8$	3					0.1406	0.1695
	4					0.2314	0.2650
	5				0.0703	0.3253	0.3619
	6				0.1479	0.4208	0.4597
	7				0.2300	0.5172	0.5579
	8				0.3151	0.6143	0.6565
	9			0.0752	0.4024	0.7118	0.7552
	10			0.1533	0.4912	0.8097	0.8542
	20			1.0117	1.4205	1.7966	1.8477
	30			1.9266	2.3795	2.7892	2.8441
	40			2.8643	3.3501	3.7839	3.8415
	50		0.7693	3.8148	4.3271	4.7798	4.8395

Table 4: Exact values of $\overline{L}_0$

	Q	$\nu = 0.01$	$\nu = 0.1$	$\nu = 0.5$	$\nu = 1$	$\nu = 5$	$\nu = 10$
$\lambda = 0.1$	3		0.6599	0.9859	1.0402	1.0875	1.0937
	4		1.0663	1.4662	1.5306	1.5857	1.5928
	5		1.4984	1.9535	2.0246	2.0846	2.0923
	6		1.9462	2.4448	2.5206	2.5838	2.5919
	7		2.4048	2.9384	3.0176	3.0833	3.0916
	8		2.8713	3.4336	3.5154	3.5829	3.5914
	9		3.3438	3.9299	4.0137	4.0825	4.0913
	10		3.8210	4.4269	4.5123	4.5823	4.5911
	20	3.8589	8.7117	9.4135	9.5062	9.5811	9.5906
	30	7.9904	13.6745	14.4090	14.5041	14.5808	14.5904
	40	12.3791	18.6559	19.4067	19.5031	19.5806	19.5903
	50	16.9288	23.6447	24.4054	24.5025	24.5805	24.5902
$\lambda = 1/3$	3			1.0490	1.1761	1.2983	1.3156
	4			1.4923	1.6465	1.7924	1.8126
	5			1.9513	2.1256	2.2883	2.3106
	6		1.3709	2.4194	2.6099	2.7854	2.8092
	7		1.7618	2.8936	3.0974	3.2831	3.3080
	8		2.1668	3.3723	3.5874	3.7813	3.8071
	9		2.5825	3.8543	4.0792	4.2799	4.3065
	10		3.0065	4.3389	4.5721	4.7786	4.8058
	20		7.5007	9.2580	9.5374	9.7728	9.8030
	30		12.2247	14.2278	14.5250	14.7708	14.8020
	40		17.0571	19.2125	19.5187	19.7697	19.8015
	50		21.9489	24.2033	24.5150	24.7691	24.8012
$\lambda = 2/3$	3				1.4295	1.6119	1.6386
	4				1.8871	2.1029	2.1341
	5			2.1104	2.3572	2.5968	2.6310
	6			2.5666	2.8342	3.0921	3.1287
	7			3.0307	3.3155	3.5884	3.6269
	8			3.5001	3.7997	4.0854	4.1254
	9			3.9734	4.2861	4.5827	4.6241
	10			4.4497	4.7741	5.0805	5.1230
	20			9.2941	9.6982	10.0667	10.1163
	30		11.5873	14.2036	14.6565	15.0594	15.1127
	40		16.2671	19.1416	19.6289	20.0548	20.1104
	50		21.0032	24.0958	24.6090	25.0514	25.1088
$\lambda = 0.8$	3					1.7406	1.7695
	4					2.2314	2.2650
	5				2.4703	2.7253	2.7619
	6				2.9479	3.2208	3.2597
	7				3.4300	3.7172	3.7579
	8				3.9151	4.2143	4.2565
	9			4.0752	4.4024	4.7118	4.7552
	10			4.5533	4.8912	5.2097	5.2542
	20			9.4117	9.8205	10.1966	10.2477
	30			14.3266	14.7795	15.1892	15.2441
	40			19.2643	19.7501	20.1839	20.2415
	50		21.1693	24.2148	24.7271	25.1798	25.2395

Table 5: Exact values of $\overline{I}$

	Q	$\nu = 0.01$	$\nu = 0.1$	$\nu = 0.5$	$\nu = 1$	$\nu = 5$	$\nu = 10$
$\lambda = 0.1$	3		5.3144	0.3446	0.1897	0.1208	0.1156
	4		3.5526	0.2846	0.1698	0.1184	0.1145
	5		2.7190	0.2495	0.1580	0.1169	0.1138
	6		2.2284	0.2264	0.1502	0.1160	0.1134
	7		1.9020	0.2099	0.1446	0.1153	0.1130
	8		1.6672	0.1975	0.1404	0.1148	0.1128
	9		1.4891	0.1879	0.1372	0.1144	0.1126
	10		1.3488	0.1803	0.1346	0.1140	0.1125
	20	79.5132	0.7284	0.1457	0.1228	0.1126	0.1118
	30	41.6055	0.5226	0.1342	0.1189	0.1121	0.1116
	40	29.2726	0.4198	0.1284	0.1170	0.1118	0.1115
	50	22.9177	0.3580	0.1249	0.1158	0.1117	0.1114
$\lambda = 1/3$	3			2.0098	0.9610	0.5576	0.5269
	4			1.5077	0.8301	0.5429	0.5201
	5			1.2616	0.7581	0.5342	0.5160
	6		38.2704	1.1150	0.7125	0.5284	0.5134
	7		19.6082	1.0175	0.6808	0.5243	0.5114
	8		13.4159	0.9476	0.6575	0.5213	0.5100
	9		10.3296	0.8949	0.6396	0.5189	0.5089
	10		8.4804	0.8537	0.6255	0.5170	0.5080
	20		3.4993	0.6751	0.5625	0.5085	0.5040
	30		2.4318	0.6167	0.5417	0.5057	0.5027
	40		1.9398	0.5875	0.5313	0.5043	0.5020
	50		1.6504	0.5700	0.5250	0.5034	0.5016
$\lambda = 2/3$	3				9.3781	2.4844	2.2197
	4				5.6693	2.3495	2.1618
	5			18.9377	4.4428	2.2735	2.1281
	6			10.4668	3.8327	2.2247	2.1060
	7			7.6529	3.4679	2.1907	2.0904
	8			6.2498	3.2251	2.1656	2.0788
	9			5.4097	3.0520	2.1464	2.0699
	10			4.8503	2.9222	2.1312	2.0628
	20			3.1022	2.4170	2.0645	2.0311
	30		19.4047	2.6940	2.2713	2.0428	2.0207
	40		11.0994	2.5097	2.2017	2.032	2.0155
	50		8.2968	2.4040	2.1607	2.0256	2.0124
$\lambda = 0.8$	3					5.8694	4.7802
	4					5.2839	4.5632
	5				26.4487	4.9778	4.4407
	6				15.2136	4.7896	4.3619
	7				11.4738	4.6622	4.3071
	8				9.6061	4.5702	4.2667
	9			53.0261	8.4865	4.5007	4.2357
	10			28.5335	7.7407	4.4463	4.2111
	20			8.1402	5.4123	4.2141	4.1035
	30			6.2849	4.8745	4.1410	4.0686
	40			5.5862	4.6347	4.1051	4.0513
	50		45.4535	5.2184	4.4988	4.0838	4.0409

Table 6: Exact values of $\overline{W}$

Table 7: Relative deviations from $\overline{L}_0$ in percent

$\lambda = 0.1$	$\nu = 0.01$			$\nu = 0.1$			$\nu = 0.5$			$\nu = 1$			$\nu = 5$			$\nu = 10$		
Q	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
3				23.22	-1.43	-3.53	6.03	-0.20	-0.18	3.15	-0.07	-0.05	0.65	-0.01	0.00	0.33	0.00	0.00
4				19.44	2.37	-2.25	4.83	0.28	-0.12	2.49	0.11	-0.03	0.51	0.01	0.00	0.25	0.01	0.00
5				16.82	3.84	-0.83	4.02	0.44	-0.04	2.05	0.16	-0.01	0.41	0.02	0.00	0.21	0.01	0.00
6				14.85	4.42	0.12	3.43	0.48	0.01	1.74	0.18	0.00	0.35	0.02	0.00	0.17	0.01	0.00
7				13.30	4.58	0.71	2.99	0.48	0.04	1.51	0.18	0.01	0.30	0.02	0.00	0.15	0.01	0.00
8				12.03	4.56	1.07	2.64	0.46	0.05	1.33	0.17	0.01	0.26	0.02	0.00	0.13	0.01	0.00
9				10.96	4.43	1.27	2.37	0.44	0.05	1.19	0.16	0.01	0.24	0.02	0.00	0.12	0.01	0.00
10				10.06	4.26	1.38	2.14	0.41	0.06	1.08	0.15	0.02	0.22	0.02	0.00	0.11	0.01	0.00
20	35.28	28.84	28.84	5.39	2.69	2.69	1.09	0.24	0.24	0.55	0.09	0.09	0.11	0.01	0.01	0.05	0.00	0.00
30	27.58	23.59	23.59	3.65	1.89	1.89	0.73	0.17	0.17	0.37	0.06	0.06	0.07	0.01	0.01	0.04	0.00	0.00
40	22.71	19.86	19.86	2.75	1.45	1.45	0.55	0.13	0.13	0.28	0.05	0.05	0.06	0.01	0.01	0.03	0.00	0.00
50	19.27	17.06	17.06	2.21	1.17	1.17	0.44	0.11	0.11	0.22	0.04	0.04	0.04	0.00	0.00	0.02	0.00	0.00

$\lambda = 1/3$																		
Q	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
3							16.24	-1.64	-1.86	9.06	-0.86	-0.62	2.02	-0.15	-0.07	1.03	-0.07	-0.03
4							13.81	1.16	-1.67	7.61	0.44	-0.61	1.66	0.05	-0.08	0.84	0.02	-0.04
5							12.14	2.38	-0.95	6.61	1.00	-0.37	1.43	0.14	-0.05	0.72	0.06	-0.02
6				36.03	16.03	0.34	10.87	2.95	-0.31	5.86	1.26	-0.15	1.25	0.18	-0.02	0.63	0.08	-0.01
7				33.38	16.91	3.24	9.87	3.21	0.15	5.27	1.38	0.03	1.11	0.20	0.00	0.56	0.09	0.00
8				31.22	17.26	5.24	9.05	3.31	0.48	4.80	1.41	0.15	1.00	0.21	0.02	0.51	0.10	0.01
9				29.41	17.32	6.63	8.36	3.31	0.71	4.39	1.41	0.24	0.91	0.21	0.03	0.46	0.09	0.01
10				27.86	17.20	7.60	7.76	3.27	0.87	4.06	1.39	0.30	0.84	0.20	0.03	0.42	0.09	0.02
20				18.73	13.98	9.41	4.49	2.36	1.07	2.28	0.98	0.35	0.46	0.15	0.04	0.23	0.07	0.03
30				14.13	11.13	11.13	3.13	1.73	1.73	1.57	0.72	0.72	0.32	0.11	0.11	0.16	0.05	0.05
40				11.26	9.08	9.08	2.39	1.36	1.36	1.20	0.56	0.56	0.24	0.08	0.08	0.12	0.04	0.04
50				9.30	7.59	7.59	1.93	1.11	1.11	0.97	0.46	0.46	0.19	0.07	0.07	0.10	0.03	0.03

$\lambda = 2/3$																		
Q	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
3										15.52	-2.25	-2.13	3.71	-0.52	-0.24	1.91	-0.26	-0.11
4										13.41	0.81	-2.31	3.15	0.03	-0.34	1.62	0.00	-0.16
5							20.78	5.20	-3.13	11.98	2.25	-1.62	2.78	0.30	-0.28	1.42	0.14	-0.13
6							19.07	6.54	-1.39	10.92	3.00	-0.88	2.50	0.45	-0.18	1.28	0.21	-0.09
7							17.73	7.27	-0.02	10.09	3.42	-0.26	2.29	0.54	-0.09	1.16	0.25	-0.05
8							16.64	7.66	1.02	9.42	3.66	0.22	2.11	0.59	-0.02	1.08	0.28	-0.01
9			15.72	7.88	1.81	12.6	8.85	3.79	0.60	6.7	1.98	0.62	0.05	1.3	1.00	0.29	0.02	0.7
10			14.94	7.97	2.41	12.2	8.36	3.85	0.89	6.5	1.85	0.64	0.10	1.3	0.94	0.31	0.04	0.6
20			10.43	7.18	4.27	9.2	5.63	3.47	1.84	4.7	1.19	0.60	0.27	0.9	0.60	0.29	0.13	0.4
30				28.48	24.64	24.64	8.23	6.13	6.13	4.35	2.94	2.94	0.91	0.51	0.51	0.45	0.24	0.24
40				24.98	22.20	22.20	6.85	5.30	5.30	3.57	2.52	2.52	0.73	0.44	0.44	0.37	0.21	0.21
50				22.40	20.24	20.24	5.88	4.66	4.66	3.04	2.21	2.21	0.62	0.38	0.38	0.31	0.18	0.18

$\lambda = 0.8$																		
Q	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
3													4.33	-0.70	-0.33	2.24	-0.35	-0.15
4													3.71	0.00	-0.49	1.91	-0.02	-0.23
5										13.85	2.75	-2.20	3.29	0.36	-0.41	1.69	0.16	-0.20
6										12.71	3.69	-1.23	2.98	0.56	-0.28	1.53	0.26	-0.14
7										11.80	4.23	-0.42	2.75	0.68	-0.16	1.40	0.32	-0.08
8										11.08	4.54	0.23	2.55	0.76	-0.05	1.30	0.36	-0.03
9							18.25	9.57	2.29	10.46	4.72	0.74	2.40	0.80	0.04	1.22	0.38	0.01
10							17.42	9.72	3.07	9.94	4.83	1.14	2.26	0.83	0.11	1.15	0.39	0.05
20							12.68	9.10	5.65	7.00	4.58	2.56	1.53	0.82	0.40	0.77	0.40	0.19
30							10.39	8.07	8.07	5.62	4.03	4.03	1.20	0.73	0.73	0.60	0.36	0.36
40							8.93	7.22	7.22	4.77	3.60	3.60	1.01	0.66	0.66	0.51	0.32	0.32
50				27.39	25.08	25.08	7.90	6.55	6.55	4.18	3.25	3.25	0.87	0.60	0.60	0.44	0.29	0.29

Table 8: Relative deviations from $\overline{I}$ in percent

$\lambda = 0.1$

Q	$\nu = 0.01$ C1	C2	C3	$\nu = 0.1$ C1	C2	C3	$\nu = 0.5$ C1	C2	C3	$\nu = 1$ C1	C2	C3	$\nu = 5$ C1	C2	C3	$\nu = 10$ C1	C2	C3
3				16.18	-0.99	-2.46	4.81	-0.16	-0.15	2.54	-0.06	-0.04	0.54	0.00	0.00	0.27	0.00	0.00
4				14.88	1.82	-1.72	4.01	0.23	-0.10	2.08	0.09	-0.03	0.43	0.01	0.00	0.21	0.01	0.00
5				13.45	3.07	-0.67	3.40	0.37	-0.03	1.75	0.14	0.00	0.35	0.01	0.00	0.18	0.00	0.00
6				12.18	3.62	0.10	2.94	0.41	0.01	1.50	0.15	0.00	0.31	0.02	0.00	0.15	0.01	0.00
7				11.09	3.82	0.59	2.58	0.41	0.03	1.31	0.15	0.01	0.26	0.02	0.00	0.13	0.01	0.00
8				10.14	3.84	0.90	2.30	0.40	0.04	1.16	0.15	0.01	0.23	0.02	0.00	0.12	0.01	0.00
9				9.32	3.77	1.08	2.07	0.38	0.05	1.04	0.14	0.01	0.21	0.02	0.00	0.11	0.01	0.00
10				8.61	3.65	1.19	1.88	0.36	0.05	0.95	0.13	0.01	0.19	0.02	0.00	0.10	0.01	0.00
20	25.68	20.99	20.99	4.74	2.37	2.37	0.97	0.22	0.22	0.49	0.08	0.08	0.10	0.01	0.01	0.05	0.00	0.00
30	22.23	19.02	19.02	3.23	1.68	1.68	0.65	0.15	0.15	0.33	0.06	0.06	0.07	0.01	0.01	0.03	0.00	0.00
40	18.95	16.57	16.57	2.45	1.29	1.29	0.49	0.12	0.12	0.25	0.04	0.04	0.05	0.01	0.01	0.02	0.00	0.00
50	16.37	14.49	14.49	1.97	1.05	1.05	0.40	0.10	0.10	0.20	0.03	0.03	0.04	0.00	0.00	0.02	0.00	0.00

$\lambda = 1/3$

Q	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
3							5.92	-0.60	-0.68	3.92	-0.37	-0.27	0.98	-0.07	-0.03	0.51	-0.04	-0.01
4							6.10	0.52	-0.73	3.76	0.22	-0.30	0.89	0.03	-0.04	0.45	0.01	-0.02
5							5.91	1.16	-0.46	3.50	0.53	-0.20	0.80	0.08	-0.03	0.41	0.03	-0.01
6				5.36	2.39	0.05	5.63	1.53	-0.16	3.24	0.70	-0.08	0.73	0.10	-0.01	0.37	0.05	-0.01
7				8.11	4.11	0.79	5.32	1.73	0.08	3.00	0.78	0.02	0.66	0.12	0.00	0.33	0.05	0.00
8				9.61	5.31	1.61	5.02	1.84	0.27	2.79	0.82	0.09	0.61	0.12	0.01	0.31	0.06	0.01
9				10.43	6.14	2.35	4.74	1.88	0.40	2.60	0.84	0.14	0.56	0.13	0.02	0.28	0.06	0.01
10				10.87	6.71	2.97	4.48	1.89	0.50	2.43	0.83	0.18	0.52	0.13	0.02	0.26	0.06	0.01
20				9.99	7.46	5.02	2.79	1.47	0.67	1.44	0.62	0.22	0.30	0.09	0.02	0.15	0.04	0.02
30				8.16	6.43	6.43	1.99	1.10	1.10	1.01	0.46	0.46	0.21	0.07	0.07	0.10	0.03	0.03
40				6.75	5.44	5.44	1.54	0.87	0.87	0.78	0.37	0.37	0.16	0.05	0.05	0.08	0.03	0.03
50				5.70	4.65	4.65	1.25	0.72	0.72	0.63	0.30	0.30	0.13	0.05	0.05	0.06	0.02	0.02

$\lambda = 2/3$

Q	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
3										1.04	-0.15	-0.14	0.64	-0.09	-0.04	0.36	-0.05	-0.02
4										1.57	0.09	-0.27	0.65	0.01	-0.07	0.35	0.00	-0.04
5							1.09	0.27	-0.16	1.82	0.34	-0.24	0.64	0.07	-0.06	0.34	0.03	-0.03
6							1.73	0.59	-0.13	1.93	0.53	-0.16	0.61	0.11	-0.04	0.32	0.05	-0.02
7							2.13	0.87	0.00	1.98	0.67	-0.05	0.59	0.14	-0.02	0.31	0.07	-0.01
8							2.38	1.10	0.15	1.98	0.77	0.05	0.56	0.16	0.00	0.29	0.08	0.00
9							2.53	1.27	0.29	1.97	0.84	0.13	0.54	0.17	0.01	0.28	0.08	0.00
10							2.63	1.40	0.42	1.94	0.89	0.21	0.52	0.18	0.03	0.27	0.09	0.01
20							2.57	1.77	1.05	1.57	0.97	0.51	0.36	0.18	0.08	0.18	0.09	0.04
30				3.08	2.67	2.67	2.24	1.67	1.67	1.28	0.87	0.87	0.28	0.16	0.16	0.14	0.08	0.08
40				3.99	3.55	3.55	1.96	1.52	1.52	1.08	0.77	0.77	0.23	0.14	0.14	0.12	0.07	0.07
50				4.27	3.86	3.86	1.73	1.37	1.37	0.94	0.68	0.68	0.20	0.12	0.12	0.10	0.06	0.06

$\lambda = 0.8$

Q	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
3													0.35	-0.06	-0.03	0.21	-0.03	-0.01
4													0.38	0.00	-0.05	0.22	0.00	-0.03
5										0.39	0.08	-0.06	0.39	0.04	-0.05	0.22	0.02	-0.03
6										0.64	0.18	-0.06	0.39	0.07	-0.04	0.22	0.04	-0.02
7										0.79	0.28	-0.03	0.38	0.09	-0.02	0.21	0.05	-0.01
8										0.89	0.37	0.02	0.37	0.11	-0.01	0.20	0.05	0.00
9							0.34	0.18	0.04	0.96	0.43	0.07	0.36	0.12	0.01	0.19	0.06	0.00
10							0.59	0.33	0.10	1.00	0.48	0.11	0.35	0.13	0.02	0.19	0.06	0.01
20							1.36	0.98	0.61	1.01	0.66	0.37	0.27	0.14	0.07	0.14	0.07	0.03
30							1.40	1.09	1.09	0.91	0.65	0.65	0.22	0.13	0.13	0.11	0.07	0.07
40							1.33	1.07	1.07	0.81	0.61	0.61	0.19	0.12	0.12	0.10	0.06	0.06
50				1.00	0.91	0.91	1.24	1.03	1.03	0.73	0.57	0.57	0.17	0.11	0.11	0.08	0.06	0.06

Table 9: Relative deviations from $\overline{W}$ in percent

$\lambda = 0.1$

	$\nu = 0.01$			$\nu = 0.1$			$\nu = 0.5$			$\nu = 1$			$\nu = 5$			$\nu = 10$		
Q	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
3				-11.82	0.73	1.80	-11.01	0.37	0.34	-5.37	0.1	0.08	-0.36	0.01	0.00	-0.10	0.00	0.00
4				-17.18	-2.10	1.99	-12.15	-0.70	0.29	-5.36	-0.2	0.07	-0.32	-0.01	0.00	-0.09	-0.01	0.00
5				-21.19	-4.84	1.05	-12.39	-1.34	0.11	-5.09	-0.4	0.02	-0.28	-0.01	0.00	-0.07	0.00	0.00
6				-24.18	-7.19	-0.20	-12.22	-1.71	-0.04	-4.74	-0.5	-0.01	-0.26	-0.02	0.00	-0.06	0.00	0.00
7				-26.45	-9.12	-1.41	-11.86	-1.89	-0.14	-4.40	-0.5	-0.03	-0.22	-0.01	0.00	-0.05	0.00	0.00
8				-28.17	-10.68	-2.49	-11.41	-1.98	-0.20	-4.09	-0.5	-0.04	-0.20	-0.01	0.00	-0.05	0.00	0.00
9				-29.49	-11.92	-3.42	-10.94	-2.02	-0.24	-3.81	-0.5	-0.04	-0.19	-0.02	0.00	-0.05	-0.01	0.00
10				-30.49	-12.93	-4.20	-10.48	-2.02	-0.28	-3.56	-0.5	-0.05	-0.17	-0.01	0.00	-0.05	-0.01	0.00
20	-15.58	-12.73	-12.73	-33.38	-16.65	-16.65	-7.06	-1.58	-1.58	-2.11	-0.3	-0.34	-0.09	-0.01	-0.01	-0.02	0.00	0.00
30	-25.11	-21.48	-21.48	-32.54	-16.87	-16.87	-5.25	-1.23	-1.23	-1.49	-0.2	-0.25	-0.06	0.00	0.00	-0.02	0.00	0.00
40	-30.82	-26.95	-26.95	-31.08	-16.38	-16.38	-4.17	-0.99	-0.99	-1.15	-0.2	-0.20	-0.05	-0.01	-0.01	-0.01	0.00	0.00
50	-34.55	-30.59	-30.59	-29.54	-15.71	-15.71	-3.45	-0.83	-0.83	-0.93	-0.2	-0.16	-0.04	0.00	0.00	-0.01	0.00	0.00

$\lambda = 1/3$

Q	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
3							-4.63	0.47	0.53	-2.88	0.27	0.20	-0.24	0.02	0.01	-0.06	0.00	0.00
4							-6.04	-0.51	0.73	-3.19	-0.18	0.26	-0.23	-0.01	0.01	-0.06	0.00	0.00
5							-6.87	-1.35	0.53	-3.27	-0.49	0.18	-0.21	-0.02	0.01	-0.05	0.00	0.00
6				-2.88	-1.28	-0.03	-7.32	-1.99	0.21	-3.24	-0.70	0.08	-0.19	-0.03	0.00	-0.05	-0.01	0.00
7				-5.47	-2.77	-0.53	-7.57	-2.46	-0.12	-3.15	-0.82	-0.02	-0.18	-0.03	0.00	-0.05	-0.01	0.00
8				-7.76	-4.29	-1.30	-7.66	-2.80	-0.41	-3.04	-0.90	-0.10	-0.17	-0.03	0.00	-0.04	-0.01	0.00
9				-9.78	-5.76	-2.20	-7.66	-3.04	-0.65	-2.93	-0.94	-0.16	-0.16	-0.04	-0.01	-0.04	-0.01	0.00
10				-11.56	-7.14	-3.16	-7.59	-3.20	-0.85	-2.81	-0.96	-0.20	-0.14	-0.03	-0.01	-0.04	-0.01	0.00
20				-21.41	-15.98	-10.75	-6.05	-3.18	-1.44	-1.88	-0.81	-0.29	-0.09	-0.03	-0.01	-0.02	-0.01	0.00
30				-24.61	-19.38	-19.38	-4.75	-2.64	-2.64	-1.38	-0.63	-0.63	-0.06	-0.02	-0.02	-0.02	-0.01	-0.01
40				-25.43	-20.51	-20.51	-3.87	-2.20	-2.20	-1.09	-0.51	-0.51	-0.05	-0.02	-0.02	-0.01	0.00	0.00
50				-25.27	-20.63	-20.63	-3.26	-1.87	-1.87	-0.89	-0.43	-0.43	-0.04	-0.01	-0.01	-0.01	0.00	0.00

$\lambda = 2/3$

Q	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
3										-0.48	0.07	0.07	-0.09	0.01	0.01	-0.03	0.00	0.00
4										-0.78	-0.05	0.14	-0.10	0.00	0.01	-0.03	0.00	0.00
5							-0.73	-0.18	0.11	-0.96	-0.18	0.13	-0.10	-0.01	0.01	-0.03	0.00	0.00
6							-1.28	-0.44	0.09	-1.07	-0.29	0.09	-0.09	-0.02	0.00	-0.03	0.00	0.00
7							-1.69	-0.69	0.00	-1.13	-0.39	0.03	-0.09	-0.02	0.00	-0.02	0.00	0.00
8							-2.00	-0.92	-0.12	-1.17	-0.45	-0.03	-0.08	-0.02	0.00	-0.02	0.00	0.00
9							-2.23	-1.12	-0.26	-1.18	-0.51	-0.08	-0.08	-0.02	0.00	-0.02	-0.01	0.00
10							-2.41	-1.29	-0.39	-1.19	-0.55	-0.13	-0.08	-0.03	0.00	-0.02	-0.01	0.00
20							-2.89	-1.99	-1.18	-1.05	-0.65	-0.34	-0.05	-0.03	-0.01	-0.01	-0.01	0.00
30				-5.52	-4.78	-4.78	-2.73	-2.03	-2.03	-0.89	-0.60	-0.60	-0.04	-0.02	-0.02	-0.01	-0.01	-0.01
40				-8.78	-7.80	-7.80	-2.49	-1.93	-1.93	-0.76	-0.54	-0.54	-0.03	-0.02	-0.02	-0.01	0.00	0.00
50				-10.81	-9.76	-9.76	-2.26	-1.79	-1.79	-0.67	-0.49	-0.49	-0.03	-0.02	-0.02	-0.01	0.00	0.00

$\lambda = 0.8$

Q	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
3													-0.05	0.01	0.00	0.00	0.00	0.00
4													-0.05	0.00	0.01	0.00	0.00	0.00
5										-0.18	-0.04	0.03	-0.05	-0.01	0.01	-0.02	0.00	0.00
6										-0.31	-0.09	0.03	-0.05	-0.01	0.00	-0.01	0.00	0.00
7										-0.39	-0.14	0.01	-0.05	-0.01	0.00	-0.01	0.00	0.00
8										-0.45	-0.19	-0.01	-0.05	-0.01	0.00	-0.01	0.00	0.00
9							-0.26	-0.14	-0.03	-0.50	-0.22	-0.03	-0.05	-0.01	0.00	-0.01	0.00	0.00
10							-0.47	-0.26	-0.08	-0.53	-0.26	-0.06	-0.04	-0.02	0.00	-0.01	0.00	0.00
20							-1.31	-0.94	-0.58	-0.57	-0.37	-0.21	-0.03	-0.02	-0.01	-0.01	-0.01	0.00
30							-1.45	-1.12	-1.12	-0.53	-0.38	-0.38	-0.03	-0.02	-0.02	-0.01	-0.01	-0.01
40							-1.43	-1.16	-1.16	-0.48	-0.36	-0.36	-0.02	-0.02	-0.02	-0.01	0.00	0.00
50				-2.32	-2.12	-2.12	-1.37	-1.14	-1.14	-0.44	-0.34	-0.34	-0.02	-0.01	-0.01	0.00	0.00	0.00

	$\overline{L}_0$		$\overline{I}$	$\overline{W}$		$\overline{L}_0$		$\overline{I}$	$\overline{W}$	
Q	rel.	abs.	rel.	rel.	abs.	rel.	abs.	rel.	rel.	abs.
λ=0.1	ν= 0.01					ν = 0.1				
3						-95.47	-0.44	-66.54	48.60	2.58
4						-75.49	-0.62	-57.79	66.72	2.37
5						-58.98	-0.71	-47.17	74.27	2.02
6						-47.42	-0.76	-38.89	77.19	1.72
7						-39.19	-0.79	-32.67	77.94	1.48
8						-33.14	-0.80	-27.95	77.63	1.29
9						-28.54	-0.81	-24.27	76.78	1.14
10						-24.96	-0.82	-21.36	75.65	1.02
20	50.62	1.42	36.85	-22.4	-17.77	-10.12	-0.78	-8.90	62.63	0.46
30	1.18	0.08	0.95	-1.1	-0.45	-5.73	-0.69	-5.08	51.10	0.27
40	-4.02	-0.42	-3.35	5.5	1.60	-5.73	-0.60	-3.24	41.10	0.17
50	-4.33	-0.62	-3.68	7.8	1.78	-2.42	-0.51	-2.15	32.35	0.12
λ = 1/3	ν = 0.1									
6	848.14	1.73	126.33	-67.9	-25.98					
7	282.35	1.21	68.66	-46.3	-9.07					
8	125.90	0.84	38.74	-31.3	-4.20					
9	62.17	0.57	22.05	-20.7	-2.14					
10	31.41	0.37	12.26	-13.0	-1.11					
20	-4.70	-0.19	-2.51	5.4	0.19					
30	-0.62	-0.04	-0.36	1.1	0.03					
40	2.19	0.22	1.31	-5.0	-0.10					
50	3.91	0.53	2.40	-10.6	-0.18					
λ= 2/3	ν = 0.1					ν = 0.5				
5						1832.51	2.02	95.9	-64.1	-12.14
6						685.85	1.60	62.3	-45.9	-4.80
7						384.00	1.40	46.1	-36.5	-2.80
8						266.56	1.33	38.1	-32.0	-2.00
9						212.45	1.36	34.2	-30.2	-1.63
10						185.21	1.45	32.6	-29.9	-1.45
20						162.99	3.74	40.2	-45.2	-1.40
30	251.8	3.16	27.25	-48.8	-9.47	174.74	6.76	47.6	-57.9	-1.56
40	93.6	2.43	14.96	-32.9	-3.65	181.89	9.96	52.0	-66.1	-1.66
50	58.4	2.34	11.13	-28.2	-2.34	186.27	13.22	54.9	-71.7	-1.72
λ= 0.8	ν = 0.1					ν = 0.5				
9						5723.9	-0.08	105.6	-81.1	-43.03
10						2548.4	-0.15	85.8	-68.5	-19.5
20						431.8	-1.01	46.4	-44.7	-3.64
30						375.1	-1.93	50.4	-52.3	-3.28
40						372.7	-2.86	55.4	-59.7	-3.34
50	831	6.39	30.21	-70.3	-31.97	376.4	-3.81	59.3	-65.5	-3.42

Table 10: Rel. and abs. deviations for Sigman and Simchi-Levi's approximation, rel. deviations are given in percent

$\lambda = 0.1$	Q	$\nu = 0.01$	$\nu = 0.1$	$\nu = 0.5$	$\nu = 1$	$\nu = 5$	$\nu = 10$
	3		-3.54%	-0.18%	-0.05%	0.00%	0.00%
	4		-2.26%	-0.12%	-0.03%	0.00%	0.00%
	5		-0.83%	-0.04%	-0.01%	0.00%	0.00%
	6		0.12%	0.01%	0.00%	0.00%	0.00%
	7		0.71%	0.04%	0.01%	0.00%	0.00%
	8		1.07%	0.05%	0.01%	0.00%	0.00%
	9		1.27%	0.05%	0.01%	0.00%	0.00%
	10		1.38%	0.06%	0.02%	0.00%	0.00%
	20	28.84%	2.69%	0.24%	0.09%	0.01%	0.00%
	30	2.64%	1.89%	0.17%	0.06%	0.01%	0.00%
	40	-2.25%	1.45%	0.13%	0.05%	0.01%	0.00%
	50	-2.63%	1.17%	0.11%	0.04%	0.00%	0.00%

$\lambda = 1/3$	Q	$\nu = 0.01$	$\nu = 0.1$	$\nu = 0.5$	$\nu = 1$	$\nu = 5$	$\nu = 10$
	3			-1.87%	-0.63%	-0.06%	-0.03%
	4			-1.67%	-0.62%	-0.08%	-0.04%
	5			-0.94%	-0.37%	-0.05%	-0.02%
	6		0.34%	-0.31%	-0.15%	-0.02%	-0.01%
	7		3.24%	0.15%	0.03%	0.00%	0.00%
	8		5.23%	0.48%	0.15%	0.02%	0.01%
	9		6.63%	0.71%	0.24%	0.03%	0.01%
	10		7.60%	0.87%	0.30%	0.03%	0.02%
	20		-1.27%	1.05%	0.36%	0.05%	0.02%
	30		2.47%	1.73%	0.72%	0.11%	0.05%
	40		4.07%	1.36%	0.56%	0.08%	0.04%
	50		4.94%	1.11%	0.46%	0.07%	0.03%

$\lambda = 2/3$	Q	$\nu = 0.01$	$\nu = 0.1$	$\nu = 0.5$	$\nu = 1$	$\nu = 5$	$\nu = 10$
	3				-2.16%	-0.24%	-0.09%
	4				-2.33%	-0.35%	-0.16%
	5			-3.16%	-1.62%	-0.28%	-0.13%
	6			-1.41%	-0.89%	-0.18%	-0.09%
	7			-0.04%	-0.26%	-0.09%	-0.05%
	8			1.01%	0.23%	-0.02%	-0.01%
	9			1.81%	0.61%	0.05%	0.02%
	10			2.41%	0.89%	0.10%	0.04%
	20			4.27%	1.84%	0.27%	0.13%
	30		24.64%	6.13%	2.94%	0.51%	0.24%
	40		22.20%	5.30%	2.52%	0.44%	0.21%
	50		20.24%	4.66%	2.21%	0.38%	0.18%

$\lambda = 0.8$	Q	$\nu = 0.01$	$\nu = 0.1$	$\nu = 0.5$	$\nu = 1$	$\nu = 5$	$\nu = 10$
	3					-0.34%	-0.14%
	4					-0.48%	-0.25%
	5				-2.23%	-0.40%	-0.19%
	6				-1.20%	-0.28%	-0.14%
	7				-0.43%	-0.16%	-0.09%
	8				0.22%	-0.05%	-0.04%
	9			2.30%	0.73%	0.04%	0.01%
	10			3.07%	1.15%	0.11%	0.05%
	20			5.65%	2.56%	0.40%	0.19%
	30			8.07%	4.03%	0.73%	0.36%
	40			7.22%	3.60%	0.66%	0.32%
	50		25.09%	6.55%	3.25%	0.60%	0.29%

Table 11: Relative deviations of $\overline{L}_0^I$ from $\overline{L}_0$

$\lambda = 0.1$	Q	$\nu = 0.01$	$\nu = 0.1$	$\nu = 0.5$	$\nu = 1$	$\nu = 5$	$\nu = 10$
	3		-2.47%	-0.14%	-0.04%	0.00%	0.00%
	4		-1.73%	-0.10%	-0.03%	0.00%	0.00%
	5		-0.67%	-0.03%	0.00%	0.00%	0.00%
	6		0.10%	0.01%	0.00%	0.00%	0.00%
	7		0.59%	0.03%	0.01%	0.00%	0.00%
	8		0.90%	0.04%	0.01%	0.00%	0.00%
	9		1.08%	0.05%	0.01%	0.00%	0.00%
	10		1.19%	0.05%	0.01%	0.00%	0.00%
	20	20.99%	2.37%	0.22%	0.08%	0.01%	0.00%
	30	2.13%	1.68%	0.15%	0.06%	0.01%	0.00%
	40	-1.87%	1.29%	0.12%	0.04%	0.01%	0.00%
	50	-2.23%	1.05%	0.10%	0.03%	0.00%	0.00%

$\lambda = 1/3$	Q	$\nu = 0.01$	$\nu = 0.1$	$\nu = 0.5$	$\nu = 1$	$\nu = 5$	$\nu = 10$
	3			-0.68%	-0.27%	-0.03%	-0.01%
	4			-0.74%	-0.31%	-0.04%	-0.02%
	5			-0.46%	-0.20%	-0.03%	-0.01%
	6		0.05%	-0.17%	-0.08%	-0.01%	-0.01%
	7		0.78%	0.08%	0.02%	0.00%	0.00%
	8		1.61%	0.27%	0.09%	0.01%	0.01%
	9		2.35%	0.40%	0.14%	0.01%	0.01%
	10		2.97%	0.50%	0.18%	0.02%	0.01%
	20		-0.68%	0.65%	0.23%	0.03%	0.01%
	30		1.42%	1.10%	0.46%	0.07%	0.03%
	40		2.44%	0.87%	0.37%	0.05%	0.03%
	50		3.03%	0.72%	0.30%	0.05%	0.02%

$\lambda = 2/3$	Q	$\nu = 0.01$	$\nu = 0.1$	$\nu = 0.5$	$\nu = 1$	$\nu = 5$	$\nu = 10$
	3				-0.14%	-0.04%	-0.02%
	4				-0.27%	-0.07%	-0.03%
	5			-0.17%	-0.25%	-0.06%	-0.03%
	6			-0.13%	-0.15%	-0.05%	-0.02%
	7			0.00%	-0.05%	-0.02%	-0.01%
	8			0.14%	0.05%	0.00%	0.00%
	9			0.29%	0.13%	0.01%	0.00%
	10			0.42%	0.21%	0.03%	0.01%
	20			1.05%	0.51%	0.08%	0.04%
	30		2.67%	1.67%	0.87%	0.16%	0.08%
	40		3.55%	1.52%	0.77%	0.14%	0.07%
	50		3.86%	1.37%	0.68%	0.12%	0.06%

$\lambda = 0.8$	Q	$\nu = 0.01$	$\nu = 0.1$	$\nu = 0.5$	$\nu = 1$	$\nu = 5$	$\nu = 10$
	3					-0.03%	-0.01%
	4					-0.05%	-0.03%
	5				-0.06%	-0.05%	-0.02%
	6				-0.06%	-0.04%	-0.02%
	7				-0.03%	-0.02%	-0.01%
	8				0.02%	-0.01%	-0.01%
	9			0.04%	0.07%	0.01%	0.00%
	10			0.10%	0.12%	0.02%	0.01%
	20			0.61%	0.37%	0.07%	0.03%
	30			1.09%	0.65%	0.13%	0.07%
	40			1.07%	0.61%	0.12%	0.06%
	50		0.91%	1.03%	0.57%	0.11%	0.06%

Table 12: Relative deviations of $\overline{I}^{I}$ from $\overline{I}$

able aggregate steady state probabilities for this process. These probabilities express invariance properties of the system similar to those of Theorem 3.2 and its Corollary 3.3.

Theorem 4.1 *For the stable queueing-inventory process Z_e under $(0,Q)$*

$$\mathsf{P}_e(Y=0)=\frac{\lambda}{Q\nu}, \qquad \textit{and} \qquad \mathsf{P}_e(X>0,Y=k)=\frac{\lambda}{Q\mu} \quad \textit{for } k=1,...,Q.$$

Corollary 4.2 *For the stable queueing-inventory process Z_e under $(0,Q)$*

$$\mathsf{P}_e(X=0)=1-\lambda\left(\frac{1}{Q\nu}+\frac{1}{\mu}\right)+\pi_e(0,0) \quad \textit{and} \quad \mathsf{P}_e(X>0,Y>0)=\frac{\lambda}{\mu}. \tag{24}$$

The proof of Theorem 4.1 is along the same lines as the proof of Theorem 3.2. The explicit results allow to compare performance characteristics of the systems. We reduce the problem of computing explicit expressions to that of computing boundary probabilities similar to the case of $\Pi(1)$ in Section 3.3. The boundary probabilities can be determined in a similar way.

Corollary 4.3 *For the stable queueing-inventory process Z_e under $(0,Q)$*

$$\mathsf{P}_e(X=0)-\mathsf{P}_l(X=0)=\pi_e(0,0), \quad \mathsf{P}_l(Y=0)-\mathsf{P}_e(Y=0)=\pi_l(0,0).$$

Theorem 4.4 *The difference between the performance measures under $(0,Q)$ respectively under $\Pi(1)$ are*

$$\begin{array}{ll} \overline{I}_e-\overline{I}_l=1-\lambda(\frac{1}{Q\nu}+\frac{1}{\mu}), & \overline{W}_e-\overline{W}_l=-\frac{1}{Q\nu}+\frac{\pi_e(0,0)}{\lambda}, \\ \alpha_{2,e}-\alpha_{2,l}=\pi_l(0,0), & \beta_e-\beta_l=\frac{\nu}{\lambda}\pi_e(0,0). \end{array}$$

$\lambda_{R,e}$ and $\lambda_{R,l}$ coincide.

Proof. The proof of Theorem 4.4 is presented in Appendix A.4, p. 31. □

Corollary 4.5 *The total costs as in (15) of the system according under $\Pi(1)$ are lower then those of the system under $(0,Q)$ if and only if*

$$\frac{\lambda h}{v_w}>\frac{\mathsf{P}_e(X>0,Y=0)}{\mathsf{P}_e(X=0,Y>0)}.$$

Proof of Corollary 4.5. From Theorem 4.4 and Theorem 4.1 we deduce that

$$\begin{aligned} \overline{I}_e-\overline{I}_l &= 1-\lambda\left(\frac{1}{Q\nu}+\frac{1}{\mu}\right)=\mathsf{P}_e(X=0,Y>0) \\ \text{and} \qquad \overline{W}_e-\overline{W}_l &= -\frac{1}{Q\nu}+\frac{\pi_e(0,0)}{\lambda}=-\frac{1}{\lambda}\mathsf{P}_e(X>0,Y=0). \end{aligned}$$

Hence, we obtain

$$TC_e-TC_l=\mathsf{P}_e(X=0,Y>0)\cdot h-\frac{1}{\lambda}\mathsf{P}_e(X>0,Y=0)\cdot v_w.$$

The result follows immediately. □

Remark 4.6 *For asymptotically vanishing replenishment lead time ($\nu\to\infty$), the optimal order quantity is the classical economic order quantity $Q^*=\sqrt{\dfrac{2A\lambda}{h}}$.*

5 Discussion and conclusion

Construction and evaluation of models for the interaction of service systems and attached inventories is a challenging field in Operations Research. We contributed in this paper by investigating the interplay of well established models in both fields:
The $M/M/1/\infty$ queue and the inventory under (r, Q) policy. We proved that an analytical Markovian performance analysis of the model is possible to a certain extent, especially with respect to stability analysis. But there remains a residual part of the fundamental steady state behaviour that is not accessible by this approach. We developed an algorithm taylored to this system to solve this problem numerically. In course of this developement we reuse an approximation by Sigman and Simchi-Levi [SSL92] which is adapted to the system under light traffic and combine this with a new heuristics into an interpolation algorithm. The extensive numerical experiments show that the proposed interpolation algorithm works well and is fast.

We have worked out in detail the case of $(0, Q)$ policy with threshold 1 for the queue length. The principles of the procedure can be applied to the (r, Q) policy case as well, and moreover to the case of (r, S) policy. We further believe that a similar approach can be applied to many other complex systems where the steady state probabilities are not explicitly at hand.

It should be noted that Berman and Kim [BK01], [BK04], have proved that the (r, Q) is only suboptimal. They proved that optimal policies have to take into consideration as decision variable the queue length and that optimal policies are of monotone structure. So their findings point the way to further research. We expect that an interplay of explicit functional relations and adapted numerical procedures as developed in this paper will be a promising path to further results.

An interesting further problem is performance assessment for queues with finite waiting room and associated inventory as it is dealt with for the case of lost sales in [Sch04][Section 1.3.5].

A Appendix

A.1 Numerical calculation of the stationary distribution

The last step for deriving some of the performance measures in Theorem 3.6 is to calculate the queue-length-zero boundary probabilities $\pi_0 = (\pi(0,1), ..., \pi(0, Q-1))$, which can be done by solving the complete steady state equations numerically. We review methods to calculate the equilibrium π of the system. Several methods rely on Z being a QBD process, where *matrix geometric methods* (pioneered by Marcel Neuts [Neu78], [Neu81]) are often well suited. Latouche et al. [LPT98] showed that for an irreducible continuous time QBD process there always exists a matrix-geometric invariant measure of the form $\mathbf{x}_n = \mathbf{x}_0 \mathbf{R}^n$, $n \in \mathbb{N}$. The rate matrix $\mathbf{R}$ is the minimal non negative solution to a quadratic equation in $\mathbf{R}$

$$\mathbf{A}_0 + \mathbf{R}\mathbf{A}_1 + \mathbf{R}^2\mathbf{A}_2 = \mathbf{0}, \tag{25}$$

where $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2$ are defined in the proof of Proposition 2.5. The vector $\mathbf{x}_0$ satisfies $\mathbf{x}_0[\mathbf{A}_1^{(0)} + \mathbf{R}\mathbf{A}_2] = 0$. If the QBD process is ergodic (25) can be converted to

$$\mathbf{R}_0 = 0, \quad \mathbf{R}_{i+1} = -\mathbf{A}_0\mathbf{A}_1^{-1} - \mathbf{R}_i^2\mathbf{A}_2\mathbf{A}_1^{-1},$$

and the sequence $\{\mathbf{R}_i, i \geq 0\}$ is entry-wise nondecreasing and converges monotonically to the unique nonnegative matrix $\mathbf{R}$ which has a spectral radius less than 1 and satisfies (25). A similar way to compute $\mathbf{R}$ is described by Lee and Seo [LS98]. The *logarithmic reduction algorithm* of Latouche and Ramaswami [LR93] has quadratic convergence. The algorithm was speeded up by Naoumov et al. [NKW96] using matrix factorization. A numerically more stable implementation of the logarithmic reduction algorithm has been proposed by Ye [Ye02].

Other methods calculate π use transform techniques. These reduce the infinite-state problem to a finite set of linear equations involving the usually unknown vector π_0. This linear equation can be solved with the help of the singularities of the coefficient matrix of a vector generating function. This approach is used in Iravani and Posner [IP95]. An application of this method to the system under consideration can be found in Schwarz [Sch04][Section 1.5.2.2]. The same method of solution results when applying spectral analysis [GHT96] to the continuous time case.

Bini and Meini [BM96] developed a quadratically convergent *cyclic reduction algorithm* for $M/G/1$-type matrices. Applied to QBD processes cyclic reduction leads to an algorithm similar to the logarithmic reduction algorithm proposed by Latouche and Ramaswami [LR93], which is slightly faster. Bini and Meini [BM97] further improved this method and achieved lower computational cost along with higher numerical stability.

Akar and Sohraby [AS97] proposed an *invariant subspace approach* for the numerical solution of $M/G/1$ and $G/M/1$ type Markov chains, which both include QBD chains as special cases.

Using a different form for the level probabilities π_n, $n = 0, 1, ...$ Chakka and Mitrani [CM93] developed a computational method called *spectral expansion*. The invariant probability vector of the process is expressed in terms of eigenvalues and left eigenvectors of a matrix polynomial.

Different methods are compared by Tran and Do [TD00] and Mitrani and Chakka [CM93].

A.2 Proof of Theorem 3.8

Ergodicity holds in both systems for the same parameter values, see Proposition 2.5. From the PASTA property the virtual and actual waiting times are the same in both systems. Furthermore, the actual waiting time is a function of an enlarged arrival disposition for customers which contains all the service times of the customers in queue and the intermediate replenishment lead times necessary to serve the customers in line.

Let W_n denote the waiting time associated with customer n in Sigman and Simchi-Levi's system and let $\widetilde{W}_n$ denote the waiting time associated with customer n in our original system. S_n is the service time of customer n and T_n the inter-arrival time between customer n and customer $n+1$. R_n denotes the replenishment time of the n-th order. S_n, T_n and $R_n, n \in \mathbb{N}$, are each independent identically distributed (i.i.d.) sequences of random variables which obey an exponential distribution with parameter μ, λ and ν, respectively. Furthermore these sequences are independent of one another. Therefore the sequences $(W_n : n \in \mathbf{N})$ and $(\widetilde{W}_n : n \in \mathbf{N})$ constitute irreducible and positively recurrent Markov Chains which are not time-homogeneous.

Without loss of generality we assume that both systems are started totally empty at time 0. This means that there are no customers in the system and the inventory is depleted. The first customer who arrives at time T_0 is customer 1. In Sigman and Simchi-Levi's system a replenishment order is triggered after the service of the customers with number $m \cdot Q$, $m \in \mathbb{N}_0$. The next customer to be served experiences possibly the replenishment lead time as waiting time. The waiting time of the $m \cdot Q + 1$st customers is thus given by the Lindley equation

$$W_{m \cdot Q+1} = (W_{m \cdot Q} + S_{m \cdot Q} + R_{m+1} - T_{m \cdot Q})_+.$$

The waiting time of all other customers is given by

$$W_{l+1} = (W_l + S_l - T_l)_+.$$

Define $\Delta_n := S_n - T_n$ for all $n \in \mathbb{N}_0$, with $S_0 := 0$. Then the vector of the first Q waiting times denoted by $\mathbf{W}_1$ is given by

$$\mathbf{W}_1 = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_Q \end{pmatrix} = \begin{pmatrix} \max(0, R_1 + \Delta_0) \\ \max(0, \Delta_1, \Delta_1 + R_1 + \Delta_0) \\ \vdots \\ \max(0, \Delta_{Q-1}, \Delta_{Q-1} + \Delta_{Q-2}, .., \Delta_{Q-1} + .. + R_1 + \Delta_0) \end{pmatrix}.$$

The waiting times for the next Q customers are eventually prolonged by the next replenishment of the inventory. They are given by the vector

$$\mathbf{W}_2 = \begin{pmatrix} W_{Q+1} \\ W_{Q+2} \\ \vdots \\ W_{2Q} \end{pmatrix} = \begin{pmatrix} \max(0, W_Q + R_2 + \Delta_Q) \\ \max(0, \Delta_{Q+1}, \Delta_{Q+1} + W_Q + R_2 + \Delta_Q) \\ \vdots \\ \max(0, \Delta_{2Q-1}, \Delta_{2Q-1} + \Delta_{2Q-2}, .., \Delta_{2Q-1} + .. + W_Q + R_2 + \Delta_Q) \end{pmatrix}.$$

Similarly, we define the vector process by means of f: $\mathbf{W}_{n+1} = f(\mathbf{W}_n, \mathbf{U}_n)$, $n \in \mathbb{N}$, where

$$f : \mathbb{R}^Q \times \mathbb{R}^{Q+1} \to \mathbb{R}^Q,$$

$$\left(\begin{pmatrix} w_1 \\ \vdots \\ w_Q \end{pmatrix}, \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_Q \\ r \end{pmatrix} \right) \mapsto \begin{pmatrix} \max(0, w_Q + r + \delta_1) \\ \max(0, \delta_2, \delta_2 + w_Q + r + \delta_1) \\ \vdots \\ \max(0, \delta_Q, \delta_Q + \delta_{Q-1}, .., \delta_Q + .. + w_Q + r + \delta_1) \end{pmatrix}$$

and where

$$\mathbf{U}_n = \left(\Delta_{Qn}, \Delta_{Qn+1}, ..., \Delta_{(Q+1)n-1}, R_{n+1} \right)^T, \quad n \in \mathbb{N}.$$

The vectors $(\mathbf{U}_n : n \in \mathbb{N})$ constitute an i.i.d. sequence. Hence, the sequence $(\mathbf{W}_n : n \in \mathbb{N})$ is a time-homogeneous Markov Chain. Define $h(\mathbf{W}_n) := \frac{1}{Q} \sum_{i=1}^{Q} W_{Q(n-1)+i}$, $n \in \mathbb{N}$ then

$$\frac{1}{NQ} \sum_{i=1}^{NQ} W_i = \frac{1}{N} \sum_{n=1}^{N} h(\mathbf{W}_n).$$

In the $M/M/1/\infty$-system with inventory management under $\Pi(1)$ policy a replenishment order is triggered after the customer number $m \cdot Q + 1, m \in \mathbb{N}$ enters the server to start his service. This customer does not experience the replenishment lead time as waiting time. Hence, when starting the system totally empty we obtain for our original system with late replenishment a time-homogeneous Markov Chain as follows: $\widetilde{W}_1 = 0$,

$$\widetilde{\mathbf{W}}_1 = \begin{pmatrix} \widetilde{W}_2 \\ \widetilde{W}_3 \\ \vdots \\ \widetilde{W}_{Q+1} \end{pmatrix} = \begin{pmatrix} \max(0, \Delta_1 + R_1) \\ \max(0, \Delta_2, \Delta_2 + \Delta_1 + R_1) \\ \vdots \\ \max(0, \Delta_Q, \Delta_Q + \Delta_{Q-1}, .., \Delta_Q + ... + \Delta_1 + R_1) \end{pmatrix},$$

and $\widetilde{\mathbf{W}}_{n+1} = f(\widetilde{\mathbf{W}}_n, \widetilde{\mathbf{U}}_n)$ for all $n \in \mathbb{N}$ with an i.i.d. sequence $\widetilde{\mathbf{U}}_n = (\Delta_{Qn+1}, ..., \Delta_{(Q+1)n}, R_{n+1})^T$. The Markov chains $(\mathbf{W}_n : n \in \mathbb{N})$ and $(\widetilde{\mathbf{W}}_n : n \in \mathbb{N})$ have the same transition matrix. Furthermore

$$\frac{1}{NQ+1} \sum_{i=1}^{NQ+1} \widetilde{W}_i = \frac{1}{NQ+1} \left(\widetilde{W}_1 + \sum_{i=2}^{NQ+1} \widetilde{W}_i \right) = \frac{1}{NQ+1} \widetilde{W}_1 + \frac{NQ}{NQ+1} \frac{1}{NQ} \sum_{i=2}^{NQ+1} \widetilde{W}_i.$$

Thus,

$$\lim_{N\to\infty} \frac{1}{NQ+1} \sum_{i=1}^{NQ+1} \widetilde{W}_i = \lim_{N\to\infty} \frac{1}{NQ} \sum_{i=2}^{NQ+1} \widetilde{W}_i = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} h(\widetilde{\mathbf{W}}_n).$$

From the ergodic theorem independently of the initial state
$\overline{W}^S = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} h(\mathbf{W}_n) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} h(\widetilde{\mathbf{W}}_n) = \overline{W}$ almost surely. □

A.3 Discussion of the range of applicability for Sigman and Simchi-Levi's approximation in Section 3.4.2

The indicated restriction of the light traffic approximation of Sigman and Simchi-Levi is due to the fact that Sigman and Simchi-Levi's approximation neglects the impact of Q on the behaviour of the system. If Q is small the following error arises: replenishment orders have to be made very often and if ν is small the mean time for a replenishment order to arrive might take up the largest part of the time between two orders. More customers might have accumulated than items arrived to the inventory. Consequently, a busy period must not be smaller than the mean time between two orders as assumed by this approximation. Examples where this phenomenon arises are exhibited in table 13. This underestimation of the mean waiting times results in large rel. deviations for $\overline{L}_0$. Thus, Sigman and Simchi-Levi's approximation can only be used if $\frac{1}{\nu} \ll \frac{Q}{\lambda}$, where ν and λ are small. The conclusion is not to use this approximation for small order size Q.

A.4 Omitted proofs from Section 3.3

Proof of Theorem 3.6. The average on-hand inventory position $\overline{I}$ is

$$\overline{I} = \sum_{k=0}^{Q} k \sum_{i=0}^{\infty} \pi(i,k) \stackrel{\star}{=} \sum_{k=1}^{Q} k \left(\frac{\lambda}{Q\mu} + \pi(0,k) \right) = \frac{(Q+1)\lambda}{2\mu} + \sum_{k=1}^{Q-1} k\pi(0,k),$$

Table 13: Critical examples for Sigman and Simchi-Levi's approximation

parameters			λ_R^{-1}	$\overline{W}$	$\overline{W}^S$
$\lambda = 1/3,$	$\nu = 0.1,$	$Q = 6$	18	38.27	12.29
$\lambda = 2/3,$	$\nu = 0.5,$	$Q = 5$	7.5	18.94	6.8
$\lambda = 2/3,$	$\nu = 0.5,$	$Q = 6$	9	10.47	5.67
$\lambda = 0.8,$	$\nu = 0.5,$	$Q = 9$	11.25	53.03	10
$\lambda = 0.8,$	$\nu = 0.5,$	$Q = 10$	12.5	28.53	9.03

where $\stackrel{\star}{=}$ follows from Theorem 3.2.

λ_R is determined using Palm calculus. Let the ARRIVAL OF ORDERS state pairs be

$$R = \{\big((n,0),(n,Q)\big) : n \geq 1\} \subseteq E_Z \times E_Z - \text{ diag } (E_Z \times E_Z)$$

and let N_R be the point process counting the R-transitions of Z. From formula (1.4.15) on p. 38 of Baccelli and Brémaud [BB03] the intensity of N_R is

$$\lambda_R = E[N_R((0,1])] = \sum_{((n,k),(m,l))\in R} \pi(n,k)\, q\big((n,k),(m,l)\big). \tag{26}$$

$$\text{Therefore, } \lambda_R = \sum_{((n,0),(n,Q))\in R} \pi(n,0)\, q\big((n,0),(n,Q)\big) \stackrel{\star}{=} \nu \mathsf{P}(X > 0, Y = 0) \stackrel{\star\star}{=} \frac{\lambda}{Q},$$

where $\stackrel{\star}{=}$ follows from (8) on p. 9 and $\stackrel{\star\star}{=}$ follows from Theorem 3.2. Furthermore, we have

$$\alpha_2 = \mathsf{P}(Y > 0) = 1 - \mathsf{P}(X > 0, Y = 0) - \pi(0,0) \stackrel{\star}{=} 1 - \frac{\lambda}{Q\nu} - \pi(0,0).$$

Here $\stackrel{\star}{=}$ follows from Theorem 3.2, p. 7. The calculation of β needs more theoretical background, we refer to [Sch04][Appendix A.3.3]. The calculation of the mean waiting time of the customers is done using the transform technique mentioned in Section A.1. For details see Schwarz [Sch04]. With average inventory $\overline{I}$ the mean number of customers in the system is given by

$$\overline{L}_s = \frac{\lambda}{Q\nu\overline{\pi}_0}\left(\frac{Q+1}{2} + \frac{Q\nu}{\mu} - \overline{I}\right).$$

The average number of waiting customers $\overline{L}$ and the average waiting time $\overline{W}$ are

$$\overline{L} = \sum_{k=0}^{Q}\sum_{n=1}^{\infty}(n-1)\pi(n,k) = \overline{L}_s - \mathsf{P}(X > 0) = \overline{L}_s - \lambda\left(\frac{1}{Q\nu} + \frac{1}{\mu}\right),$$

$$\overline{W} = \frac{1}{\lambda}\overline{L} = \frac{1}{Q\nu\overline{\pi}_0}\left(\frac{Q-1}{2} + \frac{\lambda(Q\nu+\mu)^2}{Q\nu\mu^2} - \overline{I}\right).$$

Recall that the waiting time of a customer ends at the moment when he enters the server excluding possible replenishment lead time from the customer waiting time. □

Proof of Theorem 4.4.

$$\overline{I}_e = \sum_{k=1}^{Q} k \sum_{n=0}^{\infty}\pi_e(n,k) = \sum_{k=1}^{Q} k\mathsf{P}_e(X > 0, Y = k) + \sum_{k=1}^{Q} k\pi_e(0,k) \stackrel{\star}{=} \frac{(Q+1)\lambda}{2\mu} + \sum_{k=1}^{Q} k\pi_e(0,k),$$

where $\stackrel{\star}{=}$ follows from Theorem 4.1, p. 26.

$\lambda_{R,e}$ is calculated in the same way as $\lambda_{R,l}$ and $\lambda_{R,e} = \lambda_{R,l}$ is then obvious. $\alpha_{2,e} = \mathsf{P}_e(Y > 0) = 1 - \frac{\lambda}{Q\nu}$, from Theorem 4.1 with $Y_{net} = Y$. Hence, from (12) $\alpha_{2,e} - \alpha_{2,l} = \pi_l(0,0)$.

To compute β_e is more involved, we refer to [Sch04][Appendix A.3.3].

In order to find $\overline{W}_e$ we have to repeat the calculations using the transform technique. As is explained in [Sch04] the average number of customers in system under $(0, Q)$ policy is

$$\overline{L}_{s,e} = \frac{\lambda}{Q\nu\left(1 - \lambda\left(\frac{1}{Q\nu} + \frac{1}{\mu}\right)\right)}\left(\frac{Q+1}{2} + \frac{Q\nu}{\mu} - \overline{I}_e\right)$$

where $\overline{I}_e$ denotes the average inventory of the early replenishment system. The average number of waiting customers $\overline{L}_e$ and the average waiting time $\overline{W}_e$ are

$$\begin{aligned}
\overline{L}_e &= \overline{L}_{s,e} - \mathsf{P}_e(X > 0) = \overline{L}_{s,e} - \lambda\left(\frac{1}{Q\nu} + \frac{1}{\mu}\right) + \pi_e(0,0), \\
\overline{W}_e &= \frac{1}{\lambda}\overline{L}_e = \frac{1}{Q\nu\left(1 - \lambda\left(\frac{1}{Q\nu} + \frac{1}{\mu}\right)\right)}\left(\frac{Q-1}{2} + \frac{\lambda(Q\nu+\mu)^2}{Q\nu\mu^2} - \overline{I}_e\right) + \frac{\pi_e(0,0)}{\lambda}.
\end{aligned}$$

Recall that the waiting time does not include the waiting time for the next relenishment order to arrive if the first customer in line finds the inventory empty. Having calculated $\overline{W}_e$ we make use of Theorem 3.8 from Section 3.4.2 which states $\overline{W}_e = \overline{W}_l$ after adding the additional waiting time a customer experiences because of outstanding replenishment orders if he is first in line. We will denote this newly defined average waiting time for the early replenishment system which includes the replenishment lead time by $\overline{W}_r$ then

$$\overline{W}_r = \overline{W}_e + \frac{1}{\lambda}\mathsf{P}_e(X > 0, Y = 0).$$

This leads to the following equivalent equalities

$$\begin{aligned}
\overline{W}_l - \overline{W}_e &= \frac{\mathsf{P}_e(X > 0, Y = 0)}{\lambda} \\
\frac{1}{Q\nu(1 - \lambda(\frac{1}{Q\nu} + \frac{1}{\mu}))}\left(\overline{I}_e - \overline{I}_l\right) &= \frac{\pi_e(0,0)}{\lambda} + \frac{\mathsf{P}_e(X > 0, Y = 0)}{\lambda} - \frac{\mathsf{P}_e(Y = 0)}{\lambda}. \\
\text{Hence, from Theorem 4.1} \qquad \overline{I}_e - \overline{I}_l &= 1 - \lambda(\frac{1}{Q\nu} + \frac{1}{\mu}). \\
\text{Moreover,} \qquad \overline{W}_e - \overline{W}_l &= -\frac{\mathsf{P}_e(X > 0, Y = 0)}{\lambda} = -\frac{1}{Q\nu} + \frac{\pi_e(0,0)}{\lambda}.
\end{aligned}$$

This completes the proof. □

References

[AS97] N. Akar and K. Sohraby. An invariant subspace approach in M/G/1 and G/M/1 type Markov chains. *Comm. Statist.– Stochastic Models*, 13(3):381 – 416, 1997.

[BB03] F. Baccelli and P. Brémaud. *Elements of Queueing Theory.* Springer-Verlag, Berlin – Heidelberg – New York, 2003.

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

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

[BK04] O. Berman and E. Kim. Dynamic inventory strategies for profit maximization in a service facility with stochastic service, demand and lead time. *Mathematical Methods of Operations Research*, 60:497–521, 2004.

[BM96] D.A. Bini and B. Meini. On the solution of a nonlinear matrix equation arising in queueing problems. *SIAM J. Matrix Anal. Appl.*, 17(4):906 – 926, 1996.

[BM97] D.A. Bini and B. Meini. Improved cyclic reduction for solving queueing problems. *Numerical Algorithms*, 15:57 – 74, 1997.

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

[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.

[CM93] R. Chakka and I. Mitrani. Application and evaluation of the spectral expansion solution method. In U. Herzog, N. Götz, and M. Rettelbach, editors, *Proceedings of the QMIPS-Workshop in Formalism, Principles and State-of-the-art*, pages 253 – 267, Erlangen, 1993.

[GH85] D. Gross and C.M. Harris. *Fundamentals of Queueing Theory. Second Edition.* John Wiley and Sons, Inc., Chichester – New York – Brisbane – Toronto – Singapore, 1985.

[GHT96] H.R. Gail, S.L. Hantler, and B.A. Taylor. Spectral analysis of M/G/1 and G/M/1 type Markov chains. *Advances in Applied Probability*, 28:114 – 165, 1996.

[IP95] S.M.R. Iravani and M.J.M. Posner. An M/G/1 queue with cyclic service times. *Questa*, 21:145 – 165, 1995.

[LPT98] G. Latouche, C.E.M Pearce, and P.G. Taylor. Invariant measures for quasi-birth-and-death processes. *Comm. Statist.– Stochastic Models*, 14:443 – 460, 1998.

[LR93] G. Latouche and V. Ramaswami. A logarithmic reduction algorithm for quasi-birth-death processes. *Journal of Applied Probability*, 30:650 – 674, 1993.

[LS98] T.E. Lee and J.W. Seo. Stochastic cyclic flow lines: Non-blocking, Markovian models. *Journal of the Operations Research Society*, 49:537 – 548, 1998.

[Neu78] M.F. Neuts. Markov chains with applications in queueing theory, which have a matrix-geometric invariant vector. *Advances in Applied Probability*, 10:185 – 212, 1978.

[Neu81] M.F. Neuts. *Matrix Geometric Solutions in Stochastic Models - An Algorithmic Approach.* John Hopkins University Press, Baltimore, MD, 1981.

[NKW96] V. Naoumov, U. Krieger, and D. Wagner. Analysis of a multi-server delay-loss system with a general Markovian arrival process. In S.R. Chakravarthy and A.S. Alfa, editors, *Matrix-analytical methods in stochastic models*, volume 183 of *Lecture Note in Pure and Applied Mathematics*. Marcel Dekker, New York, 1996.

[Sch81] C. Schneeweiß. *Modellierung industrieller Lagerhaltungssysteme.* Springer-Verlag, Berlin – Heidelberg – New York, 1981.

[Sch04] M. Schwarz. *Stochastic Models in Inventory Theory.* Shaker Verlag, Aachen, 2004.

[SSD03a] M. Schwarz, C. Sauer, and H. Daduna. Exponential queueing systems with an attached inventory under (r, Q)- or (r, S)-policy. Preprint 2003-9, Schwerpunkt Mathematische Statistik und Stochastische Prozesse, Fachbereich Mathematik der Universität Hamburg, 2003.

[SSD+03b] M. Schwarz, C. Sauer, H. Daduna, R. Kulik, and R. Szekli. M/M/1-systems with different inventory management policies and lost sales. Preprint 2003-07, Schwerpunkt Mathematische Statistik und Stochastische Prozesse, Fachbereich Mathematik der Universität Hamburg, 2003.

[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.

[TD00] H.T. Tran and T.V. Do. Computational aspects for steady state analysis of QBD processes. *Periodica Polytechnica, Ser. El. Eng.*, 44(2):179 – 200, 2000.

[Ye02] Q. Ye. On Latouche-Ramaswami's logarithmic reduction algorithm for quasi-birth-and-death processes. *Comm. Statist.– Stochastic Models*, 18:449 – 467, 2002.